

Evaluation of Document Similarity between Corporate AI Ethics Guideline and UNESCO's AI Ethics Recommendation

A Scalable Approach to Analyze AI Ethics Guideline
Documents Utilizing Natural Language Processing

Ken Suzuki

Former Project Researcher, Institute for Future Initiatives, The University of Tokyo

Evaluation of Document Similarity between Corporate AI Ethics Guideline and UNESCO’s AI Ethics Recommendation

A Scalable Approach to Analyze AI Ethics Guideline Documents Utilizing Natural Language Processing

Ken Suzuki
Institute for Future Initiatives
The University of Tokyo
Tokyo, Japan
kennygrandcanyon@gmail.com

Abstract—The number of corporate guidelines and governance strategies on artificial intelligence (AI) among private sector entities has been increasing in the past five years. Corresponding to the self-regulatory behavior to achieve responsible use of AI by technology firms, numerous government and regulatory bodies have also introduced and imposed policies and legal frameworks to ensure the public is safe from harmful use of it. For example, the European Union (EU) has drafted its AI Act and the National Institute of Standards and Technology (NIST) has announced its AI Risk Management Framework.

We are approaching an era where various levels of governmental bodies are expected to enact regulations targeting AI companies. Hence, we should look into the details of the gap and similarity between regulators and regulatees for implementing AI Ethics. There have been a handful of studies that have investigated the content similarities and general trends in AI ethics guidelines among categorizing organization types. However, case studies on whether the current corporate guidelines comply with rules set by governmental bodies remain unclear. Therefore, our study analyzed the document similarity between UNESCO’s “Recommendation on the Ethics of Artificial Intelligence” and AI ethics guidelines published by 49 top global technology companies that utilize AI for their core business. This paper also serves as an early research effort for proposing a Natural Language Processing-based method to evaluate the compliance level of corporate AI ethics guidelines with regulations and standards proposed by governments.

This study reveals that the 49 corporate AI ethics guidelines generally share similar document similarity levels with UNESCO’s guidelines. The private sector tends to deprioritize the incorporating “Sustainability” aspect of AI such as SDGs, environmental concerns, as well as influence on local cultures. Meanwhile, UNESCO and the leading AI companies agree on the importance of users’ right to privacy and data protection. The above-mentioned

findings overlap with previous studies investigating the private and public sector’s shared interests and conflicts of interest. Therefore, we argue that quantitative research methods have a similar level of effectiveness in distilling the overarching trends from AI ethics guidelines compared to more time-consuming and labor-intensive qualitative methods.

Index Terms—Document Similarity, AI Ethics, AI Governance, Responsible AI, Natural Language Processing (NLP)

I. INTRODUCTION

Leading technology companies have created their own AI governance strategies to promote the concept of AI ethics as well as implement practices internally to achieve responsible management of artificial intelligence in the past five years. (Jobin et al., 2019; Schiff et al., 2021) The proactive self-regulatory behavior exhibited by the private sector is an expected outcome from international organizations, governments, and civil organizations becoming increasingly aware of alarming AI incidents as well as the general public’s growing concern with the negative social impact. The European Union (EU) is one of the governing bodies that initiated the early discussion to regulate and govern AI as a society. The recently drafted EU AI act has now passed the European Parliament and should prompt further discussion for national governments trying to lead the tech innovation landscape.

As we undergo the surge of governing bodies establishing foundational principles for the potential development of AI regulations, it is highly valuable to understand the cross-sectoral similarities and differences of interests toward responsible governance of AI. There have been several studies that have collected AI policy documents, guidelines, and statements to analyze the content trends observed in different

sectors: private, public, NGO, civil society, and academic society (Fjeld et al., 2020; Larsson, 2019; Jobin et al., 2019; Schiff et al., 2021; Zeng et al., 2019). The previous scholarship has contributed to summarizing the general characteristics of each sector and explained the reason some perspectives concerning the ethics of AI are particularly prioritized or deprioritized. Understanding the difference of interests among sectors is beneficial for multi-stakeholder cooperation aiming at the successful implementation of AI governance. This analysis should help governments, tech firms, and advocacy groups to propose pragmatic and achievable strategies to convert ethical guidelines into practices. We argue that the global affairs surrounding AI ethics are moving toward the next phase where legislative and industry authorities will commence enforcing rule frameworks. Therefore, we should evaluate the current success of the AI ethics implementation level by applying a role-based framework of regulator and regulatee: the organizations that are subject to regulations enforced by governing bodies. In other words, It is difficult to grasp how far apart our society is from an ideal level of AI governance without assessing the misaligned aspect of AI ethics between the private sector and the government. Acknowledging the gap between the two sectors is critical for regulators to guide companies to modify their approach to comply with regulations.

This study presents an experimental quantitative research method to evaluate the similarity level of the private sector's AI ethics guidelines and a singular selected government body's guideline. In other words, we propose a method to analyze the company's responsible AI strategy from a sole regulator point of view. The quantitative analysis employs a state-of-art language model, developed to comprehend longer sentences and paragraphs, to compare the guideline documents. We expect our research method to become a future-proof approach to conduct a simple analysis to measure the private sector's readiness for a proposed regulation or an enactment of a bill. For instance, our research methods could be valuable for measuring the readiness for the EU AI Act as the law could trigger AI auditing practices across the member states (Kop, 2021; Mökander et al., 2021). In this paper, we selected UNESCO as an example regulator and used their Recommendation on the Ethics of Artificial Intelligence (2021) as a document to compare with the ethics guidelines issued by leading AI-focused technology firms. Our research outcome will scrutinize which aspects of AI ethics are generally well recognized by the private sector and which principles are often discounted from their corporate AI governance.

II. AI ETHICS GUIDELINE AS AN EMERGING RESEARCH TOPIC

Various organizations have communicated their commitment and action plans to ensure the responsible use of AI through devoted guideline documents. There have been new studies investigating the content of AI ethics guidelines coinciding with thriving advocacy for responsible AI principles and

governance methods. We have identified three major themes in AI ethics guideline studies, and the below sections are the summary of each theme.

A. *Cross Industrial AI Ethics Guideline Trends*

Research focusing on identifying trends in AI ethics guidelines from various sectors emerged in the past five years. This type of research usually collects guidelines from the three major sectors: Private, Public, and Non-Governmental Organisations (NGOs). Public sector documents range from domestic government to intergovernmental organizations such as the United Nations and the Organization for Economic Cooperation and Development. The definition of NGO includes membership-driven organizations including IEEE and the World Economic Forum, and advocacy groups such as the Future of Humanity Institute and Partnership on AI. Zeng et al. (2018) conducted a study applying topic frequency analysis to grasp which core AI ethics concepts were more or less embraced in every three sectors. Contrary to Zeng's research approach, there are studies based on qualitative data analysis using human-based manual document interpretation and inductive coding methods (Jobin et al., 2019; Schiff et al., 2021) or adopting hand coding to identify core values and then apply cluster analysis (Fjeld et al., 2020) to distill the primary principles discovered among the source data. The previous literature shares similar findings in spite of its differences in research methodologies. First, the private sector tends to put emphasis on fewer numbers of principles compared to the public sector or NGOs. Corporate stakeholders are likely more interested in AI risks and issues relevant to their business operations, hence omitting coverage on sustainability, culture, and social aspects that require contextual adaptation of principles. Secondly, public organizations and NGOs hold border perspectives on AI ethics. This characteristic is often explained by their institutional responsibility often seeks to make the industry accountable for potential risks and issues for maintaining the safety of society.

B. *Limitation in Principle to Guarantee AI Ethics*

Despite seeing the rapid proliferation of AI ethics guidelines created by companies, internal codes of ethics do not guarantee principles are translated into corporate practices for ethical AI development and deployment. Recent studies have reminded us that a guideline is only a blueprint for success and part of the steps to practicing AI ethics as a society. Winfield and Jirotko (2018) state that there are three essential steps for building social trust in technology. The roadmap starts by applying ethical principles to AI to form guidelines and initiatives that will promote the general notion and perception of AI ethics. The next step is to establish standards for the industry to follow and comply with. The standards often do not have legally-binding power and neither have punishment when failing to satisfy the requirements. At this stage, we should expect corporate stakeholders to implement internal safety or quality assurance checklists to ensure their business

operations are abiding by industrial expectations. The final step to establishing public trust in AI is to install a regulatory framework to hold companies accountable for delivering safe, robust, and transparent products to the people. Mittelstadt (2019) also agrees that supplemental elements are needed for pursuing responsible AI. He introduces creative ideas that will support ethical AI governance: encouraging a bottom-up approach to AI ethics in the private sector, and making AI developers a license-required profession. Whittlestone et al. (2019) reiterate that converting principles to practice is the biggest challenge yet. Corporate guidelines often employ value-based principles and do not provide precise standards or criteria to actualize the stated ideals. Their research points out that principals have three weaknesses that create room for stakeholders to have a conflict of interests: interpretation is highly subjective, contents could be vague and ambiguous, and principles lack in considering the practical context. Larsson (2020) states that there should be multidisciplinary research on advancing AI governance from principles to process.

C. *Ideal Values for AI Ethics Guidelines*

As the development and deployment of AI are accelerating across the world, we acknowledge both positive social change and harmful social risks associated with this outburst of advanced technology. Therefore, the early research linked to AI ethics guidelines was directed to develop and propose reliable ethical frameworks to ensure to maximize the “good” of AI and minimize the “bad” aspects. Floridi et al. (2018) which organized the AI4People Scientific Committee have presented 20 detailed recommendations for society to have a healthy relationship with AI. They claim that AI-enabled technologies have the opportunity to enrich our human life, yet misuse and overuse could lead to degrading our lives adversely. Floridi and her colleagues’ recommendation is based on five core principles: beneficence, non-maleficence, autonomy, justice, and explicability (transparency and explainability), to navigate responsible use of AI through a governance framework formed by four critical action pillars: assess (technical and social risks), develop (regulatory and accountability framework), incentivize (relevant research, governance, and business efforts) and support (organizational capacity to commit to AI ethics). More recent literature (Floridi, 2020; Kazim and Koshiyama, 2021) that discussed the requirements for creating ideal AI ethics guidelines shares the approximately same values and risk-associated concerns as the earlier literature. Protection of human rights especially access to privacy, AI decision-making explainability and transparency, safety, accountability, economic and political impacts, and sustainability matters are four overarching viewpoints. Unfortunately, the current maturity level of AI ethics guidelines is not up to the above-mentioned ideal standards. Hagendorff (2020) criticizes that AI service providers are not taking ethical concerns seriously enough to fully commit to those values. Some companies consider AI ethics as a non-binding goal to improve their credibility and favorability. The private sector must make

better efforts to understand the social consequences of the irresponsible use of AI. They must also put equal effort into committing to all perspectives of ideal AI governance to mature from selective commitment to partial AI ethics issues aligned with business interests.

III. METHODOLOGY

A. *Research Question*

The core objective of this research is to quantitatively evaluate the degree of semantic similarity between the current corporate AI ethics guidelines and UNESCO’s ethics of AI guideline named “Recommendation on the Ethics of Artificial Intelligence”. By measuring the document similarities of the guidelines, we aim to examine the following three questions. First, we will examine how much the corporate AI governance strategy and philosophy is aligned with UNESCO. The research outcome would reveal the general gap in the AI ethics mindset among leading technology firms and international organizations. Secondly, this study will investigate which of the 10 AI ethics principles defined by UNESCO is more emphasized or overlooked by the private sector. Answering this question should reveal which risk factor surrounding AI is aligned between private and public organizations and vice versa. Healthy technology governance requires public-private partnership and identifying the value misalignment is a critical first step for building a robust future collaboration. The final question is exploring why some principles are prioritized and overlooked by commercial entities. We will explore empirical explanations of the trends observed from the NLP-based analysis result. By deepening our understanding of the corporate interests in AI ethics, we can propose ideas to navigate business stakeholders to gain attention to disregarded perspectives for achieving improved AI governance.

B. *Document Data Collection*

The source of data for this research is comprised of various forms of AI ethics guidelines. We have curated an AI ethics guideline dataset comprised of 49 documents. The collected AI ethics documentation is in varied forms: corporate AI ethics guideline paper, one section in the code of ethics guideline, a dedicated corporate website page stating the corporate AI policy, or a dedicated responsible use of technology section highlighting the use and development of AI exclusively.

The data collection criteria followed a structured approach influenced by the following three research. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework referenced in the paper by Jobin, Ienca, and Vayena (2019) provided an unbiased and systematic approach to collecting grey literature using the Google search engines. Upon collecting potential documents in the search result pages, we applied the inclusion and exclusion criteria method to filter the suitable documents offered by Schiff et al. (2021). The data filtering process functions as an extra layer of assurance to maintain the quality of the dataset. To

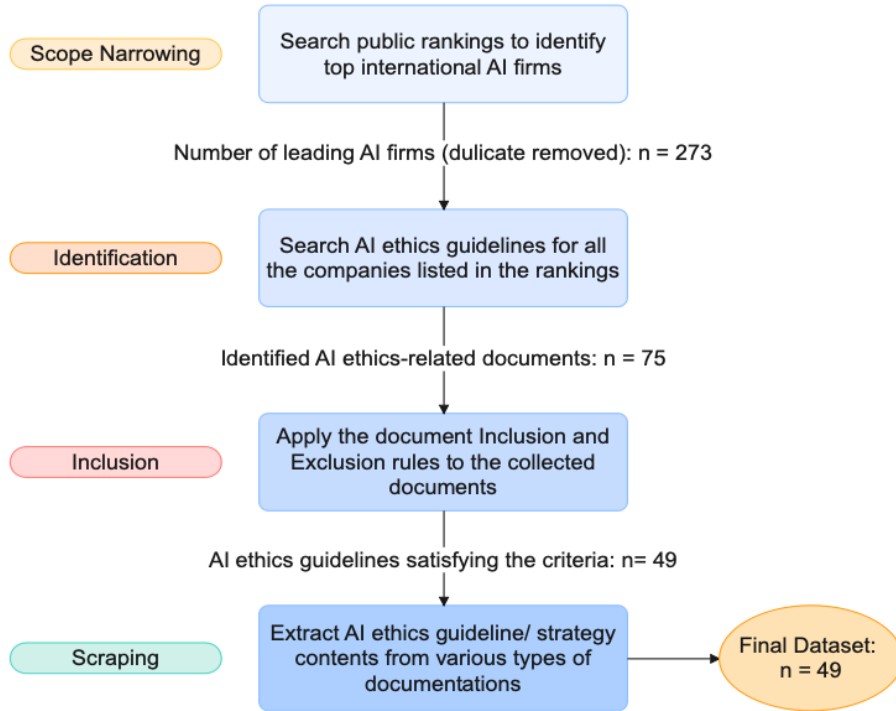


Fig. 1: Flowchart of Four-step Approach For Collecting AI Ethics Guideline Documents

determine the right search terms, Jessica et al. (2020) proposed adequate keywords to search AI ethics guidelines. Alternative words and synonyms describing “ethics” and “guideline” were applied during the data collection. Taking the previous studies into account, our document data collection became a four-step process visualized in Figure 1. First, we searched for an index and ranking of current major global tech firms as we narrowed the scope of research subject to leading AI-related service providers. We have decided to use the following five tech firm rankings referenced in Table 1. The next step is to search for AI ethics guidelines from all the firms listed in the rankings using the five search engine terms combined with a company name specified in Table 2. We used the Google Chrome search engine in a PC environment. The Incognito window feature was utilized to minimize the impact of personalized search results based on previous search histories associated with a personal Google account. To maximize the opportunity to collect the guideline documentation from each firm, we have manually scanned and verified each search result from up to the third page in Google Chrome to check if any relevant page exists within the official corporate website domain: organic results were only included and paid results were excluded. Step three is to apply the document inclusion and exclusion criteria defined in Table 3 to remove any overly broad context document, non-professional documentation such as blogs or video content, and unsuited document types including research papers, white papers, and press releases. The final process is to extract appropriate text data if the source document or

Ranking Name	Published Company	Published Date
2018 THOMSON REUTERS Top 100 Global Tech Leaders	Thomson Reuters	01-17-2018
World’s Top 25 Artificial Intelligence Companies	Algorithm-X Lab	08-20-2020
Forbes: The AI 50	Forbes	05-09-2022
Artificial Intelligence 100: Top Industry Wise Companies Listing	Analytics Insight	07-12-2022
Best AI Consultant	AIMultiple	n.d.

TABLE I: Source of AI-related Company Rankings

web page covers contents other than AI ethics guidelines. The data collection revealed that some companies tend to document their AI ethics guidelines as one of the contents in the broader Code of Ethics-related documentation available in a PDF format or create an AI ethics commitment section as a section on their official corporate website. The data collection period was from December 1st, 2022 to April 30th, 2023. To use the five-month duration effectively, there was an extensive first round of data collection for the first two months, then followed by a second round of review using the last month to check for any missing data as well as to collect new documents that were published after the first collection.

Search Engine Keywords	
1.	[company name] + AI ethics guideline
2.	[company name] + Responsible AI guideline
3.	[company name] + AI ethics policy
4.	[company name] + Responsible AI policy
5.	[company name] + Responsible AI framework

TABLE II: Corporate AI Ethics Guideline Search Terms

Rule	Inclusion	Exclusion
Publication Format	PDF, corporate webpage, corporate article page	corporate blog page, academic article (Author belongs to the firm but not representing the firm), video clip
Documentation Type	Guidelines, principles, code of ethics, commitment statement, strategy document, framework, report	Investigation report (without explicit strategies), white paper, blog post, discussion forum, event report or video
Author Type	Private entity, employee (corporate representative), committee, groups	n/a
Language	English	n/a
Availability	Publicly available through web search results	Internal corporate document available for employees only
Collection period	December 2022 to April 2023	n/a
Required Content	Clear strategies exhibiting the motivation and commitment to incorporate AI ethics principles to the firm’s operation	No strategies or action plans are articulately stated in the documentation.

TABLE III: Documentation Inclusion and Exclusion Rules

To increase the coverage of corporate AI ethics documents, this study has included non-traditional literature/ grey literature which includes extracting specific sections of a document or web page contents. We have agreed on the trade-off of the inclusion of a greyer reference with making the research outcome and analysis more inclusive and comprehensive to improve the representation of current corporate circumstances. Our new dataset is currently the most robust dataset and the exclusive corporate AI ethics guidelines. Several previous studies have a higher number of document coverage, though not have as much corporate emphasis compared to ours: Schiff et al. (2021) have collected 112 documents (private sector = 26) and Jobin, Ienca, and Vayena (2019) have collected 84 documents (private sector = 19).

C. Document Similarity Assessment

This study employed natural language processing techniques to quantitatively evaluate how much the current AI ethics guidelines contents are semantically similar to UNESCO’s recommendation document. The most common way to measure the degree of similarity or difference between two documents is a document or textual similarity assessment. The document similarity scores are produced in a numeric format ranging from 0.00 to 1.00: the higher similarity score signifies that an AI ethics document created by the specific company has a higher degree of semantic content alignment with UNESCO. The content alignment does not indicate a literal similarity but demonstrates a linguistic analogy. The language model package used for the evaluation tasks functions in a two-step procedure that will first comprehend the contents of respective documents and then produce a numeric outcome of content similarity. Figure two visually describes the assessment flow.

The document similarity was assessed from ten different aspects of AI ethics underlined in UNESCO’s Recommendation on the Ethics of Artificial Intelligence. The ten aspects are outlined in Table 4. We did not opt to conduct a raw document comparison for the following two reasons. First, comparing documents that contain several thousand words will unintentionally broaden the scope of analysis and potentially dilute the accuracy of the results. The majority of the corporate guidelines in the dataset explain their responsible use of AI approach extensively. Therefore we would like to avoid vague document comparisons by adding unnecessary complexity to the analysis. Secondly, this research required a benchmark for extracting the trends in corporate AI governance strategies. UNESCO’s ten aspects were highly useful in specifying which perspectives were aligned and which were neglected or untouched.

1) Language Model:

A state-of-art language model developed from Microsoft’s MPNet was utilized to conduct the document similarity assessment tasks. MPNet was originally introduced by Song et al. (2020) as a novel pre-training method for a robust language model that has overcome the weakness of Bidirectional Encoder Representations from Transformer (BERT) and XLNet. BERT is LM based on masked language modeling that has a tendency to ignore the dependency among predicted tokens. XLNet is the successor of BERT yet has inherent a shortcoming that cannot take the positional information into account. On the other hand, MPNet “leverages the dependency among predicted tokens through permuted language modeling (vs. MLM in BERT), and takes auxiliary position information as input to make the model see a full sentence and thus reducing the position discrepancy (vs. PLM in XLNet)” (Song et al., 2020).

The language model employed is known as all-mpnet-base-v2. The model development was led by Nils Reimers from the Ubiquitous Knowledge Processing (UKP) Lab at the Technische Universität Darmstadt (Hugging Face, 2021).

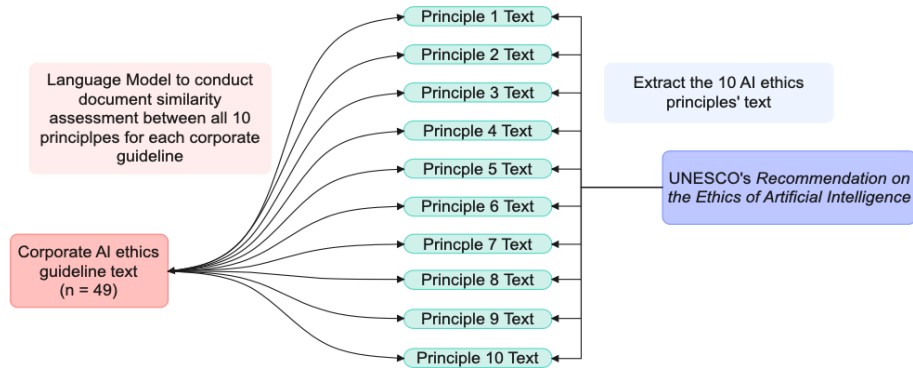


Fig. 2: Document Similarity Assessment Flow

The all-mpnet-base-v2 is a sentence transformers model fine-tuned using more than 1 billion sentence pairs dataset that has increased capacity to handle sentence and paragraph comprehension tasks.

2) UNESCO AI Ethics Guideline:

One of the core objectives of this research is to understand the similarities and differences of perspectives in the AI ethics guidelines among various private sector entities and public organizations. Therefore, we applied the following three criteria for selecting the most adequate organization representing the public sector that has already published its approach to AI ethics. 1) The organization must be serving the public good and society. 2) The organization does not represent a specific group of interests but instead serves the best interests of the general public. 3) The organization does not portray strong political beliefs or economic interests as part of its organizational mission. In light of these three criteria, we have decided to adopt the Recommendation on the Ethics of Artificial Intelligence submitted by UNESCO in November 2021. Other organizations were considered during the selection process such as the Organization for Economic Cooperation and Development (OECD): the Recommendation of the Council on Artificial Intelligence (2019), the US National Institute of Standards and Technology (NIST): Artificial Intelligence Risk Management Framework (2023), and the European Union: the EU AI Act bill (2021). UNESCO was selected due to its global representation of people and national interests while being impartial to political ideologies thanks to being one of the UN specialized agencies.

UNESCO developed its AI ethics policy as a consequence of a decision made by the member states that requested an international standard for the safe and responsible use of AI. Wong (2021) framed that the motivation behind the mandate is to draft a governance framework that is interdisciplinary and multi-stakeholder that will stimulate the proliferation of responsible use of technology among governments. Their recommendation presents core values and principles on AI ethics, how and in which policy area to apply the ethical AI manage-

ment methods, and a multi-stakeholder approach to promote and monitor the progress of AI governance involving governments, intergovernmental organizations, and civil society. The ten AI ethics principles also advocate respect, protection, and promotion of human rights and fundamental freedoms and human dignity, prospering environment and ecosystem, confirming diversity and inclusiveness, and nurturing a healthy and just society. The concept of ten principles is summarized in Table 4.

IV. RESULTS

This study compared the semantic document similarity of 49 AI ethics guidelines issued by industry-leading global technology firms and UNESCO's guidelines on AI ethics to analyze the general ethics principles converge in the private sector. The results demonstrate that corporate guidelines have approximately a 65 percent semantic document similarity rate with UNESCO's principles. Nine out of ten principles have an average similarity rate between 60 percent and 70 percent. The mean standard deviation across all the principles was 0.09, which translates to a moderate distribution of data points. The max value and minimum value for each principle have nearly a 50 percent difference. Therefore, the selected few enterprises have very high or low value and perspective alignment with UNESCO. Beyond the general trends, there were three main findings from the results each summarized in the below sections.

A. Principle 5: The most covered topic

Ensuring the right to privacy and data protection is the principle where the private sector and UNESCO appear to have the highest content alignment. Table 5 shows that Principle 5 had the highest average similarity rate among the ten principles. Table 6 supports this notion as Principle 5 has the highest frequency ($n = 23$) as the principle with the highest similarity rate. Appearing as the highest similarity principle in 23 documents is 17 documents more than Principle 9 which comes in the second place. Furthermore, no AI ethics guideline

Principles and Description	
Principle 1	Proportionality and Do No Harm (must be considered in business and social implementation of AI)
Principle 2	Safety and Security (considered AI development and training)
Principle 3	Fairness and non-discrimination (values reflected in the AI models)
Principle 4	Sustainability (to mitigate negative human, social, cultural, economic, and environmental impacts)
Principle 5	Right to Privacy, and Data Protection
Principle 6	Human oversight and determination
Principle 7	Transparency and explainability (of the AI systems and decision-making results)
Principle 8	Responsibility and accountability (through regulatory and auditing framework)
Principle 9	Awareness and literacy (to AI-related technology)
Principle 10	Multi-stakeholder and adaptive governance and collaboration

TABLE IV: 10 Principles documented in UNESCO Recommendation on the Ethics of Artificial Intelligence

had Principle 5 as the seventh, eighth, ninth, and tenth highest similarity rate ($n = 0$). Therefore, rights to privacy and data protection are highly adopted by the private sector in general, and rarely see companies excluding this principle from their guidelines.

The right to privacy and the protection and handling of personal data is often in a mutually supporting relationship to achieve both as an organization since privacy is typically infringed upon due to collecting data without user consent or utilization without prior notification. Data protection laws are currently materializing across national governments to safeguard consumers and restrict malicious or irresponsible utilization of their citizens’ data. Moreover, data-related regulation has preceded the public discussion considerably compared to AI laws and already has active regulation in selective countries. For instance, the EU’s General Data Protection Regulation (GDPR) and UK’s Data Protection Act were both enacted in 2018. The Canadian government followed suit by proposing the Consumer Privacy Protection Act (CPPA) in 2022. Therefore, recent enhancements to the legal system for protecting user privacy and data are presumably contributing to business stakeholders enforcing high internal standards regarding this subject matter to avoid future punishment due to malpractices or incidents. Not all international technology firms may not have their core operations in regulated countries.

though results suggest that the private sector as a whole is generally tackling UNESCO’s principle 5 in a preemptive fashion.

B. Principle 4: The least covered topic

Sustainability was the principle that had the lowest coverage in the corporate AI governance strategies. Principle 4 had the lowest average document similarity score of the 49 documents referring to Table 5. Uncoincidentally, we found Principle 4 dominated the count for being the principle with the worst similarity rate. 35 companies’ AI ethics framework had the lowest similarity level with Principle 4, accounting for over 70 percent of the dataset. Thus, the private sector is generally not paying as much attention to sustainability compared to the rest of the principles.

This outcome may come as a surprise since it is a regressive trend considering our society is seemingly welcoming sustainability as one of its core values. Major companies are often showing their willingness to contribute to achieving the United Nations Sustainable Development Goals (SDGs) at a surface level or at least try to communicate their business purpose reflecting sustainability values. There is a significant perception issue in sustainability that its concept is nearly interchangeable with environmental friendliness. UNESCO states that “sustainable societies relies on the achievement of a complex set of objectives on a continuum of human, social, cultural, economic and environmental dimensions” (2022, p. 21) which clearly demonstrates that sustainability demands a broader and multifaceted viewpoint. The future-proof approach to sustainability that can withstand regulations for AI service providers is to consider questions such as, how the service or product could alter human behaviors, diminish or enrich culture, or transform the current working practices and labor circumstances.

However, this underwhelming result should not concern the future of sustainability and AI. As a result of increased public engagement in sustainability, business organizations typically issue annual sustainability reports and Environmental, social, and governance (ESG) reports. We should expect the society to further develop its interests in sustainability which would consequently provoke corporate stakeholders to expand their sustainability framework to AI as well.

C. Uniform Document Similarity Level Among All Principles

Figure 3 shows the correlation rate between all potential two-principle pairings on whether the document similarity level of one principle has a negative or positive correlation with another principle. The average correlation rate of 0.87 indicates a high positive correlation rate. In other words, most AI ethics guideline documents have comparable degrees of similarity scores across the ten principles. Hence, it is rare to find a guideline document that has a significantly low or high similarity score on a particular principle. The high correlation level also suggests that principles are likely mutually dependent on embodying the foundational ethical

	Prin. 1	Prin. 2	Prin. 3	Prin. 4	Prin. 5	Prin. 6	Prin. 7	Prin. 8	Prin. 9	Prin. 10	Mean
mean	0.649370	0.640144	0.626346	0.561233	0.699394	0.628316	0.664421	0.665680	0.674071	0.617522	0.642650
std	0.090646	0.098038	0.086669	0.087055	0.100825	0.098115	0.088953	0.092453	0.097592	0.083276	0.092362
min	0.189644	0.173998	0.206271	0.200899	0.192601	0.148028	0.242611	0.220619	0.163550	0.224386	0.196261
25%	0.602693	0.610129	0.580661	0.513936	0.658497	0.579897	0.623684	0.612065	0.627909	0.584206	0.599368
50%	0.668631	0.663375	0.633357	0.569225	0.725509	0.654673	0.686528	0.695821	0.694224	0.634648	0.662599
75%	0.702077	0.702382	0.669105	0.619037	0.762905	0.693852	0.713656	0.723510	0.736288	0.666494	0.698931
max	0.771343	0.770181	0.754461	0.697254	0.813623	0.751009	0.823704	0.801979	0.799724	0.744391	0.772767

TABLE V: 10 Principles documented in UNESCO Recommendation on the Ethics of Artificial Intelligence

Ranking	Prin. 1	Prin. 2	Prin. 3	Prin. 4	Prin. 5	Prin. 6	Prin. 7	Prin. 8	Prin. 9	Prin. 10	Max. Appearance	Min. Appearance
First	2	5	2	0	23	2	4	5	6	0	Prin. 5	Prin. 4
Second	2	4	2	0	14	0	8	7	10	2	Prin. 5	Prin. 4
Third	5	5	4	0	5	0	9	10	11	0	Prin. 9	Prin. 4
Fourth	7	5	3	0	4	2	5	10	10	3	Prin. 8	Prin. 4
Fifth	12	2	3	2	1	9	6	4	9	1	Prin. 1	Prin. 5
Sixth	5	3	3	0	2	8	11	8	2	7	Prin. 7	Prin. 4
Seventh	12	2	10	2	0	11	3	4	0	5	Prin. 1	Prin. 5
Eighth	3	6	9	6	0	6	1	0	0	18	Prin. 10	Prin. 5
Ninth	1	10	12	4	0	9	2	1	1	9	Prin. 3	Prin. 5
Tenth	0	7	1	35	0	2	0	0	0	4	Prin. 4	Prin. 1

TABLE VI: 10 Principles documented in UNESCO Recommendation on the Ethics of Artificial Intelligence

values. As the corporate stakeholders deepen their understanding of regulators’ perspectives and reflect them in their internal AI governance strategies in the future, it is likely that the corporate guidelines will acquire better similarities across the board to UNESCO’s ethical ideology.

V. DISCUSSION

To expand our understanding of the current state of AI ethics guidelines created by leading AI companies, we will reflect on the findings from this research along with the learning from previous studies. Below are three prominent discussion points that I would like to introduce.

a) *Maturity of Private Sector’s Ethics Guidelines:*

An average of 65 percent document similarity level between

the private sector and UNESCO’s 10 AI ethics principles was higher than we anticipated. Previous research raised concerns about the private sector’s guideline quality for practicing AI ethics as an organization (Mittelstadt, 2019; Whittlestone et al., 2019). Hence, we expected the mean value to be lower than 50 percent which would have depicted the public-private sectoral difference of perspectives in the responsible use of AI.

On the other hand, our results reconfirmed several findings from the previous literature. We learned that the rights to privacy and data protection had high coverage in the majority of the documents, while sustainability was a neglected principle. Research by Zeng et al. (2018) has also discovered the lack of range in risk factor coverage. Companies also tend to have varied interest levels and preferences regarding ethical values:

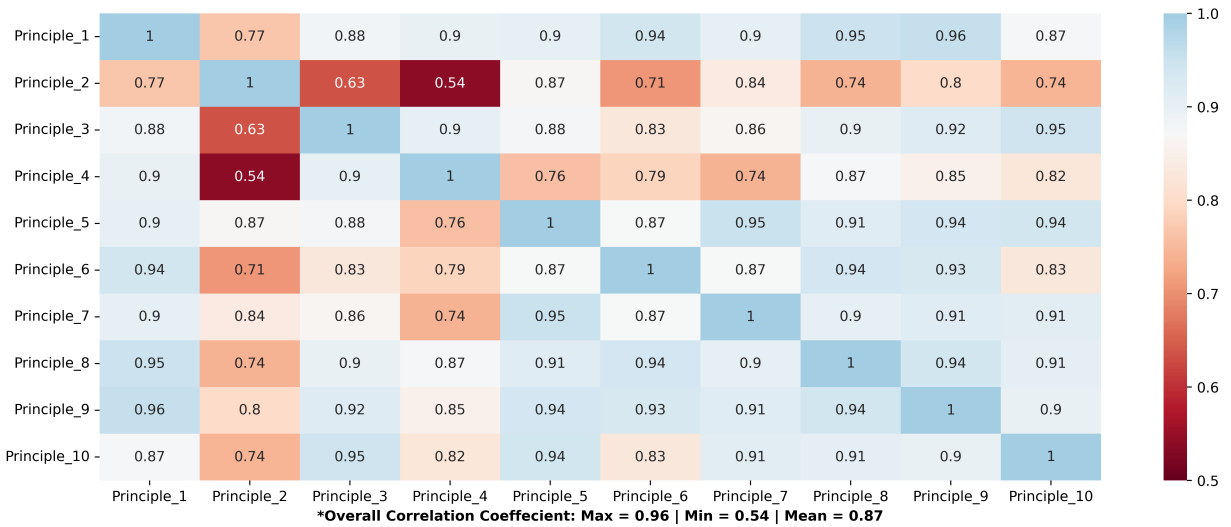


Fig. 3: Correlation Heatmap of UNESCO's 10 AI Ethics Principles

privacy rights often supersede the importance of protection of human values (Fjeld et al., 2020). Therefore, our research reinforces the notion that private entities may have a selective behavior in approaching the ethics of AI which is likely to be driven by business interests while complying with the already active data and AI regulatory frameworks.

We argue that the continuous introduction and updates of policies, frameworks, and regulations by governments and legal authorities would make a significant impact on maturing the AI governance principles issued by AI firms. The enactment of data protection acts and ongoing congressional discussion on potential data regulation bills in leading AI countries have certainly alerted the private sector to enforce higher standards of internal strategies on data management and utilization. Assuring the companies to create an AI that achieves the following is not sufficient to have a healthy relationship with AI as a society: has a non-maleficence character, has a high degree of transparency and explainability, fair algorithm decision-making, and procedures to hold companies accountable in case of unexpected incidents. Results from Jobin et al., (2019) point out that humanity-driven principles: trust, dignity, solidarity, social good, and well-being are not well addressed in guidelines across all sectors. Hence, we need regulatory bodies to initiate discussion to put a stronger spotlight on culture, humanity, and social aspects of sustainability to gain more business stakeholder interests in those matters. We expect the AI ethics guidelines to be more inclusive and holistic by fostering cross-sectoral dialogue to embrace a multi-directional approach to AI ethics.

b) *Development of Ethical Practices from Principles:*

The majority of the companies included in our research do not include clear action plans, the definition of requirements, or criteria to meet for practicing AI ethics principles. We could make excuses for the lack of enforcement strategies

as a result of an insufficient number of responsible AI practitioners available to the private sector. As we should be expecting AI regulations to be introduced in the coming years, companies should prospectively prepare to transform their AI development and deployment procedures to incorporate ethical principles.

Microsoft Corporation (2022) is one of the few firms that has already instituted requirements to achieve their responsible AI standards: accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness. Winfield and Jirotko (2018) claimed that principles and values are inadequate to gain confidence in ethical AI systems and insisted on forming standards such as precise benchmarks and system designs for companies to follow. The private sector as a whole is likely to follow the lead of front-runners in AI ethics like Microsoft since we observed low standard deviation in document similarity among the corporate AI ethics guidelines. We expect the AI ethics champions to raise the quality of corporate AI governance by raising the ceiling of a good responsible AI-practicing organization and appealing to remaining AI firms to catch up to them.

VI. CONCLUSIONS

This study has introduced a method to evaluate the AI ethics guideline similarity between the private sector and a potential regulator to measure the degree of alignment in core principle values from a regulator's perspective. The research methodology does not include manual human-based coding to identify the similarities and differences in the dataset, but instead employs an advanced language model to conduct an NLP-based document similarity assessment to produce semantic similarity scores. Our analysis indicated similar tendencies in corporate AI ethics guidelines from earlier studies. Based on UNESCO's ethical standards, the private sector had the highest level of agreement for ensuring user privacy and

data protection but had the lowest level of alignment on sustainability values.

Being able to verify similar trends in AI ethics guidelines despite the difference in research methodology raises confidence in employing an NLP-based data analysis to produce on-par quality results of qualitative methods. We advise that there are two notable benefits of employing NLP-based quantitative methods. First, the research period and human resources required could be significantly reduced. Replacing tag-based coding qualitative methods that asked for significant manual labor by multiple scholars with a machine learning-based model would result in additional benefits such as reducing the financial cost and scalability in terms of processing large numbers of complex long documents. Secondly, we should expect the accuracy of the results and findings to improve as the capability of language models advances in the future. There is a growing interest in developing large language models (LLMs) by the private sector and non-profit research institutions, hence there is a positive outlook for the future that models with higher capacity would be publicly available in the near future. Therefore, our research methodology enables regulatory agencies to evaluate the private sector's readiness for their laws or policies with fairly high accuracy, only requiring limited research scientist resources, and completing in a short time-frame.

VII. ACKNOWLEDGMENTS

I want to express our sincere gratitude to the unit and project lead, Associate Professor Arisa Ema (Institute for Future Initiatives, University of Tokyo), for her valuable guidance, feedback, and research logistics support during the research progress. Her expertise in AI ethics and governance has significantly contributed to framing the research project. I would like to express my gratitude towards two professors from the Institute for Future Initiatives, Dr. Sotaro Shibayama and Dr. Hideaki Shiroyama, who have contributed to review this paper and provided valuable feedback to improve the outcome.

I would also like to thank the NEC Corporation and their industry-academia collaboration team for funding and consulting the AI Governance collaborative research project. This research is part of the outcome of *the University of Tokyo and NEC joint research: "Research on Ethics and Legal System of AI"*.

REFERENCES

- [1] Accenture (2021). Responsible AI From Principles to Practice. <https://www.accenture.com/content/dam/accenture/final/a-com-migration/pdf/pdf-149/accenture-responsible-ai-final.pdf#zoom=50>
- [2] Adarga (2021). Responsible AI at Adarga. <https://adarga.ai/article/responsible-ai-at-adarga>
- [3] Adobe (n.d.). Adobe's Commitment to AI Ethics. <https://www.adobe.com/content/dam/cc/en/ai-ethics/pdfs/Adobe-AI-Ethics-Principles.pdf>
- [4] AIMultiple (n.d.). Best AI Consultant. <https://aimultiple.com/ai-consultant/2>
- [5] Algorithm-X Lab (2020). World's Top 25 Artificial Intelligence Companies. <https://algorithmxlab.com/blog/worlds-top-25-artificial-intelligence-companies/#Babylon>
- [6] Analytics Insight (2022). Artificial Intelligence 100: Top Industry Wise Companies Listing. <https://www.analyticsinsight.net/artificial-intelligence-100-top-industry-wise-companies-listing/>
- [7] Alphabet (n.d.). RESPONSIBILITY: Responsible AI practices. <https://ai.google/responsibility/responsible-ai-practices/>
- [8] Amazon Web Services (AWS) (n.d.). Responsible Use of Machine Learning. <https://d1.awsstatic.com/responsible-machine-learning/responsible-use-of-machine-learning-guide.pdf>
- [9] Argo AI (2022). Innovation: Quality, Continuous Improvement. <https://www.argo.ai/supplier-code-of-conduct/>
- [10] ATT (2019). Artificial Intelligence at ATT: Our Guiding Principles. https://about.att.com/innovationblog/2019/05/our_guiding_principles.html
- [11] Atos (n.d.). The Atos Blueprint for Responsible AI. <https://atos.net/en/lp/cybersecurity-magazine-ai-and-cybersecurity/the-atos-blueprint-for-responsible-ai>
- [12] Babylon (n.d.). Clinical AI Governance. <https://www.babylonhealth.com/en-gb/responsibility/clinical-ai-governance>
- [13] Baidu (n.d.). Responsible AI. https://esg.baidu.com/en/article/Responsible_AI
- [14] BMW Group (2020). BMW Group code of ethics for artificial intelligence. https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf
- [15] Boston Consulting Group (2022). AI Code of Conduct. <https://media-publications.bcg.com/AI-Code-of-Conduct.pdf>
- [16] Capgemini (n.d.). Our Code of Ethics for AI. https://www.capgemini.com/wp-content/uploads/2021/03/Capgemini_Code_of_Ethics_for_AI_2021_EN.pdf
- [17] Cisco (2022). The Cisco Responsible AI Framework. https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-framework.pdf
- [18] Cohere (2022). Responsibility: Developing safer language models. <https://cohere.com/responsibility>
- [19] Dataiku (n.d.). Corporate Social Responsibility Policy. <https://www.dataiku.com/corporate-social-responsibility-policy/>
- [20] DataRobot (n.d.). DataRobot's Framework for Trusted AI. <https://www.datarobot.com/platform/trusted-ai/>
- [21] DeeperInsights (n.d.). ETHICAL AI. <https://deeperinsights.com/our-principles/#:~:text=Ethical%20AI,lives%20for%20the%20greater%20good.>

- [22] Deepmind (n.d.) Exploring the real-world impacts of AI. <https://www.deepmind.com/about/ethics-and-society>
- [23] Deloitte (2020). Trustworthy AI™ Bridging the ethics gap surrounding AI. <https://www2.deloitte.com/us/en/pages/deloitte-analytics/solutions/ethics-of-ai-framework.html>
- [24] Deutsche Telekom (2018). AI Guidelines. <https://www.telekom.com/resource/blob/532446/f32ea4f5726ff3ed3902e97dd945fa14/dl-180710-ki-leitlinien-en-data.pdf>
- [25] European Commission (2021). Regulation Of The European Parliament And Of The Council, Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>
- [26] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1).
- [27] Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and machines*, 28, 689-707.
- [28] Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1(6), 261-262.
- [29] Forbes. (2022). Forbes: The AI 50. <https://www.forbes.com/lists/ai50/?sh=217687aa290f>
- [30] Fujifilm (2022). Fujifilm Group AI Policy. <https://holdings.fujifilm.com/en/sustainability/vision/policy/ai>
- [31] Fujitsu (2019). Fujitsu Group AI Commitment. https://www.fujitsu.com/global/documents/about/csr/humanrights/fujitsu-group-ai-commitment-201903_en.pdf
- [32] Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1), 99-120.
- [33] Hewlett Packard Enterprise (n.d.). HPE AI ETHICS AND PRINCIPLES. <https://www.hpe.com/us/en/solutions/artificial-intelligence/ethics.htmlf>
- [34] HP (n.d.). HPE AI ETHICS AND PRINCIPLES. <https://www.hpe.com/us/en/solutions/artificial-intelligence/ethics.htmlf>
- [35] Hugging Face (2021). Model Card: all-mpnet-base-v2. <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>
- [36] IBM (2020). IBM's Principles of Trust and Transparency. <https://www.ibm.com/policy/wp-content/uploads/2022/04/IBM-2020-Public-Policy-and-Legislative-Priorities-2.pdf>
- [37] Infosys (n.d.) Responsible AI. <https://www.infosys.com/content/dam/infosys-web/en/techcompass/responsible-ai.html>
- [38] Intel (n.d.). Responsible AI Pillars. <https://www.intel.com/content/www/us/en/artificial-intelligence/responsible-ai.html>
- [39] Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- [40] Kazim, E., Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9), 100314.
- [41] Kop, M. (2021). EU artificial intelligence act: the European approach to AI. Stanford-Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue.
- [42] Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*, 7(3), 437-451.
- [43] Leidos (n.d.). Trusted AI/ML. <https://www.leidos.com/enabling-technologies/artificial-intelligence-machine-learning>
- [44] Mökander, J., Axente, M., Casolari, F., Floridi, L. (2022). Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds and Machines*, 32(2), 241-268.
- [45] Meta (2021). Facebook's five pillars of Responsible AI. <https://ai.facebook.com/blog/facebook-five-pillars-of-responsible-ai/>
- [46] Microsoft (2022). Microsoft Responsible AI Standard, v2. <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmF1>
- [47] Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature machine intelligence*, 1(11), 501-507.
- [48] Motorola Solutions (2022). The Responsible Use Of Artificial Intelligence Positioning and Messaging. <https://www.motorolasolutions.com/content/dam/msi/docs/about-us/cr/the-responsible-use-of-artificial-intelligence-cr-report.pdf>
- [49] Moveworks (n.d.). Our Approach to Responsible AI. <https://www.moveworks.com/hubfs/docs/Datasheets/Moveworks-Datasheet-ResponsibleAI.pdf>
- [50] NEC (2019). NEC Group AI and Human Rights Principles. <https://www.nec.com/en/press/201904/images/0201-01-01.pdf>
- [51] NIST (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>
- [52] NTT Data Group (2019). NTT DATA Group's AI Guidelines. <https://www.nttdata.com/global/en/about-us/ai-guidelines>
- [53] Nokia (n.d.). 6 Pillars of Responsible AI. <https://www.bell-labs.com/research-innovation/ai-software-systems/responsible-ai/#gref>
- [54] NXP Semiconductors (2020). We Are NXP — A Framework for Ethical AI. <https://www.nxp.com/company/blog/we-are-nxp-a-framework-for-ethical-ai:BL-FRAMEWORK-FOR-ETHICAL-AI>
- [55] OECD (2019). Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- [56] Oki Group (2019). OKI Group AI Principles. <https://www.oki.com/en/otr/2019/n234/pdf/otr-234-R04.pdf>
- [57] PricewaterhouseCoopers International (PwC) (n.d.). PwC's Responsible AI: AI you can trust. <https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/pwc-responsible-ai.pdf>
- [58] Salesforce (n.d.). Trusted AI Principles. <https://www.salesforceairesearch.com/trusted-ai>
- [59] Samsung (2022). AI Ethics Principles. https://www.samsungsds.com/en/digital_responsibility/ai_ethics.html
- [60] SAP (2022). SAP Global AI Ethics Policy. <https://www.sap.com/documents/2022/01/a8431b91-117e-0010-bca6-c68f7e60039b.html>
- [61] Schiff, D., Borenstein, J., Biddle, J., Laas, K. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society*, 2(1), 31-42.
- [62] SenseTime (2022). AI Governance for Balanced Development. <https://oss.sensetime.com/20221014/3ec8d5aaf00d134573f1e0d214ad9270/AI%20Governance%20for%20Balanced%20Development%20White%20Paper-eng-final.pdf>
- [63] Softbank (2022). SoftBank AI Ethics Policy. <https://www.softbank.jp/en/corp/aboutus/governance/ai-ethics/>

- [64] Song, K., Tan, X., Qin, T., Lu, J., Liu, T. Y. (2020). Mpnnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33, 16857-16867.
- [65] Sony Group (2021). Sony Group AI Ethics Guidelines. https://www.sony.com/en/SonyInfo/sony_ai/responsible_ai.html
- [66] Telefónica (2018). Telefónica’s Approach to the Responsible Use of AI. <https://www.telefonica.com/es/wp-content/uploads/sites/4/2021/06/ia-responsible-governance.pdf>
- [67] Thomson Reuters. (2022). 2018 THOMSON REUTERS Top 100 Global Tech Leaders. <https://www.thomsonreuters.com/en/products-services/technology/top-100.html>
- [68] Tieto (2017). Tieto’s AI ethics guidelines. <https://www.tietoevry.com/contentassets/b097de43d84d4c84832f1fff2cb6a30d/tieto-s-ai-ethics-guidelines.pdf>
- [69] UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- [70] Vodaphone Group (n.d.). Artificial Intelligence Framework. https://www.vodafone.com/sites/default/files/2020-09/55678_Vodafone_AI_Framework_AW1_V3.pdf
- [71] Whittlestone, J., Nyrupe, R., Alexandrova, A., Cave, S. (2019). 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* (pp. 195-200).
- [72] Winfield, A. F., Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085.
- [73] Wong, A. (2021). Ethics and regulation of artificial intelligence. In *Artificial Intelligence for Knowledge Management: 8th IFIP WG 12.6 International Workshop, AI4KM 2021, Held at IJCAI 2020, Yokohama, Japan, January 7–8, 2021, Revised Selected Papers 8* (pp. 1-18). Springer International Publishing.
- [74] Workday (2019). Workday’s Commitments to Ethical AI. <https://blog.workday.com/en-us/2019/workdays-commitments-to-ethical-ai.html>
- [75] Zeng, Y., Lu, E., Huangfu, C. (2018). Linking artificial intelligence principles. arXiv preprint arXiv:1812.04814.