A Comparative Study on Ethics Guidelines for Artificial Intelligence Across Nations

Tony Szu-Hsien Lee, Shiang-Yao Liu, Yin-Ling Wei, and Li-Yun Chang(✉)

National Taiwan Normal University, Taipei 10610, Taiwan (R.O.C.)
liyunchang@ntnu.edu.tw

Abstract. This study aimed to investigate the commonality and differences among AI research and development (R&D) guidelines across nations. Content analysis was conducted on AI R&D guidelines issued by more economically developed countries because they may guide the trend of AI-based applications in education. Specifically, this study consisted of three phases: 1) information retrieval, (2) key term extraction, and (3) data visualization. First, Fisher's exact test was employed to ensure that different AI R&D guidelines (e.g., the latest ones in the US, EU, Japan, Mainland, and Taiwan) were comparable. Second, the Key Word Extraction System was developed to retrieve essential information in the guidelines. Third, data visualization techniques were performed on key terms across multiple guidelines. A word cloud revealed the similarity among guidelines (e.g., key terms that these guidelines share in common) while a color-coding scheme showed the differences (e.g., occurrence of a key term across guidelines and its frequency within a guideline). Importantly, three key terms, namely, AI, human, and development, are identified as essential commonality across guidelines. As for key terms that only extracted from particular guidelines, interestingly, results with the color-coding scheme suggested that these key terms were weighted differently depends on the developmental emphasis of a nation. Collectively, we discussed how these findings concerning ethics guidelines may shed light on AI research and development to educational technology.

Keywords: Artificial intelligence · Data visualization technique · Education · Ethics guidelines · Text mining

1 Introduction and Related Work

The rapid advances in research and development (R&D) of artificial intelligence (AI) have yielded a number of ethics guidelines. These guidelines provide guidance for new AI technologies and applications and thus are important references for developing educational technology. Since Aiken and Epstein [1] initiated a conversation concerning what is desirable and what is not in using AI in education, over the past 20 years, the growing concerns in discussing ethical issues in AI (e.g., privacy, responsibility, autonomy, justice, transparency, and beneficence) highlighted the importance of AI ethics.

T.-C. Huang et al. (Eds.): ICITL 2020, LNCS 12555, pp. 289–295, 2020.
https://doi.org/10.1007/978-3-030-63885-6_33

These ethics guidelines may vary across nations, which in turn, may influence the application of ethical principles in different fields such as industry, governments, and academia [2]. Although there are a few studies [3, 4] which compares different ethical guidelines across various stakeholders (e.g., policymakers, AI developers, key user groups or general users, educators and professionals), several critical issues exist. First, while Jobin et al.'s analysis is comprehensive, their contribution is merely descriptive [5], rather than normative. Second, while Zeng et al. attempted to use visualization techniques to explicitly establish the links among AI ethics guidelines, there approach of choosing keywords is manually-chosen, rather than data-driven. Third, the literary genre of various ethics guidelines is often neglected in previous comparative analyses. For instance, *Ethically Aligned Design* which released by IEEE [5] is 294-page long whereas *The Japanese Society for Artificial Intelligence Ethical Guidelines* [6] only has three pages. Without considering the length of content, in comparing ethics guidelines across different stakeholders, the results may be misleading.

To address the above issues, we focus on AI R&D guidelines issues by governments because these normative, official AI ethics guidelines play prominent roles in developing and implementing AI technologies. Moreover, we deliberately choose guidelines from more economically developed countries [4] given their leading statues in educational technology worldwide. Note that the length of these guidelines would be comparable. Furthermore, we utilize text mining and data visualization techniques to analyze the content. By adopting the more objective approach and by keeping in mind that we do not aim at a full analysis of all AI ethics documents, the goal of this study is to investigate the commonality and differences among these AI R&D guidelines.

Particularly, we raise the following research questions:

1. What is essential commonality across AI ethics guidelines in more economically developed countries chosen in this study?
2. What are the differences among these guidelines and how do they potentially relate to the developmental emphasis of different nations?

2 Research Method

We adopted content analysis and data visualization techniques to investigate commonality and difference of key terms among AI principles issued by governments. In particular, we focused on AI R&D principles considered by more economically development countries because they would guide the development of AI-based products.

2.1 Materials

AI R&D principles issued by the USA, EU, Japan, China, and Taiwan were selected as target content for analysis. The first three, representing more economically developed countries, together accounted for nearly half of all ethical AI principles, according to Jobin et al. [4]. The later two, China and Taiwan, were selected based on our research interest.

Below are brief sketches of these AI R&D principles, beginning from the latest one:

1. Guidance for Regulation of Artificial Intelligence Applications [7]: 10 principles from the USA.
2. Guidelines for Artificial Intelligence Technology Research and Development [8]: 8 principles and 3 core values from Taiwan.
3. Guidance for Research and Development of Artificial Intelligence [9]: 7 principles form Mainland China.
4. Ethical Principles and Democratic Prerequisites to form a responsible AI [10], 9 principles from EU.
5. The Japanese Society for Artificial Intelligence Ethical Guidelines [6]: 9 principles from Japan.

2.2 Instruments

The Key Term Extraction [11], a multilingual keyword extraction system for suggesting key terms from digital documents (PRC Patent No: ZL 00 1 22602.9.), was adopted for our content analysis. This research tool features in automatic keyword extraction, a fundamental technology in advance information retrieval system.

With a larger corpus, the precision rate of the Tseng's [12] keyword extraction algorithm is over 96% for news and over 90% for bibliographic materials, suggesting that its system quality is reliable. Moreover, this system affords both quantitative and graphical representations for the results. Resulting key terms would be ranked based on their frequency (by token) in the document, while the semantic relationship between these key terms would be shown by a key-term graph. This system is available via: http://rsp.itc.ntnu.edu.tw/SAMtool/SegWord_CGI.html.

2.3 Procedure

The procedure consisted of three phases: (1) information retrieval, (2) key term extraction, and (3) data visualization. In the phase one (information retrieval), for each guideline, number of principles and total length of principles were retrieved for examining whether they differ significantly across five guidelines. The Fisher's exact test was performed, respectively. Results showed that neither the length ($p \geq 0.05$) nor the number of principles ($p \geq 0.05$) differ across guidelines, suggesting the five guidelines were comparable. In the phase two (key term extraction), each guideline was processed by the Key Term Extraction [12]. All automatically-generated key-term graphs and the key terms were saved in a cloud drive (https://parg.co/bGGc). We examined each term carefully and kept content words, excluding function words, for further exploration. Finally, given that there were commonality and differences existed in the key terms from five guidelines, data visualization technique was adopted in the phase three.

3 Results and Discussion

We presented the commonality and differences of key terms across five guidelines by word cloud, an intuitive visualization technique to give our readers a glance into the most frequent words, and a color-coded table to show a more in-depth analysis on the coverage of different guidelines on various key terms.

3.1 Word Cloud as a Starting Point for Deeper Guideline Analyses

To derive an intuition of what information these guidelines may contain, we did the keyword summaries by HTML5 Word Cloud. Figure 1 showed the most frequent words of as a weighted list of key words. The top 5 frequent words (and times) were: AI (70), human (32), develop (24), society (18), and system (18). We also did the same analyses on separate guidelines (see cloud drive). Although font sizes of the words indicated their occurrence frequency in the guidelines, other properties did not encode specific information. Thus, in the following, we focused on the results of key term extraction.

Fig. 1. Word cloud visualization of five guidelines on the topic "AI R&D".

3.2 Color-Coded Analysis Showing Commonality and Differences in Key Terms

The key term extraction was run on individual guideline. To reveal commonality and differences of the resulting five lists of key terms, we applied a color-coding scheme (see Fig. 2).

Nation / Key term	Japan (2017)	EU (2018)	Beijing (2019)	Taiwan (2019)	USA (2020)	Cumulative frequency
AI	18	10	16	19	26	89
Human	2	9	7	8	4	30
Development	4	2	6	2	3	17
Data	0	6	3	5	2	16
Safety	3	5	2	0	3	13
Systems	0	7	7	0	3	17
Society	9	0	2	4	0	15
Information	3	2	0	0	8	13
Autonomous	0	8	0	2	2	12
Decisions	0	2	0	9	0	11
Ensure	0	3	5	0	3	11
Impact	2	0	3	0	3	8
Privacy	0	2	0	2	4	8
Fair	2	0	0	2	2	6
Risks	0	0	4	0	10	14
Benefit	0	0	2	0	9	11
Potential	0	0	4	0	7	11
Application	0	2	0	0	8	10
Nature	0	0	3	0	5	8
Respect	4	3	0	0	0	7
Security	0	2	0	0	5	7
Considered	0	0	3	0	3	6
Protection	0	4	0	0	2	6
Processes	0	2	0	0	4	6
Research	3	0	2	0	0	5
Responsibility	2	3	0	0	0	5
Rights	0	0	3	2	0	5
Integrity	2	0	0	0	3	5
Humanity	3	0	2	0	0	5
Human dignity	0	2	0	2	0	4
Environment	0	2	2	0	0	4
Implementation	0	0	2	0	2	4

Fig. 2. The distribution of key terms retrieving from AI R&D guidelines across nations. (Color figure online)

First, key terms were categorized into five colored sections to denote how common they were across five guidelines; for instance, green denotes a key term that was

mentioned in five guidelines (green = 5, blue = 4, red = 3, and 2 = yellow; for a key term that only appeared in one guideline, see cloud drive). Also, darker color denotes more frequent that the key term was mentioned.

Second, within each colored section, key terms were sequenced based on their cumulative frequency. For example, for three key terms in the green section, while they were all mentioned across five guidelines, they were listed as follows based on weights: AI(89), Human(30), and Development(17).

4 Conclusions and Future Work

This study adopted content analysis and data visualization to investigate the commonality and differences among AI R&D guidelines across nations (i.e., the US, EU, Japan, Mainland, and Taiwan). Three key terms, AI, human, and development, are identified as essential commonality across guidelines. As for key terms that only extracted from particular guidelines (e.g., risk, benefit, responsibility, rights and more), they were weighted differently in the color-coding scheme. The findings echoed prior research which suggested that AI ethics guidelines may vary across nations and cultures [2, 4], with supportive evidence from a more objective, data-driven approach. This approach could be applied to guidelines that released by other stakeholders (e.g., AI developers, key user groups or general users, educators and professionals), letting the conversation [1] moves on.

Acknowledgements. This work was financially supported by the grant MOST- 109-2634-F-003-008 from Ministry of Science and Technology of Taiwan.

References

1. Aiken, R.M., Epstein, R.G.: Ethical guidelines for AI in education: starting a conversation. Int. J. Artif. Intell. Educ. **11**, 163–176 (2020)
2. Liu, S.Y., Chang, C.H., Chao, E., Chang, L.Y., Lee, S.H.: Challenges and reflections on ethics of artificial intelligence: a literature review. J. Inf. Soc. (in Press)
3. Zeng, Y., Lu, E., Huangfu, C.: Linking artificial intelligence principles. In: Espinoza, H., Héigeartaigh, S.Ó., Huang, X., Hernández-Orallo, J., Castillo-Effen, M. (eds.) Proceedings of the AAAI Workshop on Artificial Intelligence Safety, vol. 2301, pp. 103–106. Creative Commons CC0, Honolulu (2019)
4. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**, 389–399 (2019)
5. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design (EAD1e) (2019). https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=PR&utm_source=Web&utm_campaign=EAD1e&utm_content=geias&utm_term=undefined
6. Japanese Society for Artificial Intelligence: The Japanese Society for Artificial Intelligence Ethical Guidelines (2017). http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf

7. Memorandum for the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Applications (2020). https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf
8. Digitimes: MOST announces AI R&D guidelines (2019). https://www.digitimes.com/news/a20190923PD209.html
9. Beijing Academy of Artificial Intelligence: Beijing AI Principles (2019), https://www.baai.ac.cn/news/beijing-ai-principles-en.html
10. European Commission: Artificial intelligence: Commission takes forward its work on ethics guidelines (2019). https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1893
11. Tseng, Y.H.: Multilingual keyword extraction for term suggestion. In: Proceedings of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 377–378. Association for Computing Machinery, New York (1998)
12. Tseng, Y.H.: Automatic key feature extraction from digital documents, PRC Patent No: ZL 00 1 22602.9. Application date: August 4, 2000. Effective from April 26, 2006 to August 4 (2020)