Exploring AI Ethics Syllabi Through NLP Cluster Analysis

Kerrie Hooper, Stephanie Lunn
Florida International University Miami, FL
khooper@fiu.edu, sjlunn@fiu.edu

Abstract

With new technology come new responsibilities. Examining artificial intelligence (AI) and its applications through an ethical lens has become increasingly important. Academia can play a critical role in shaping graduates who may work in both the ethical and technical spheres. To better understand how this may be integrated into higher education institutions, we assessed the content covered on AI ethics using Natural Language Processing (NLP) syllabi. A total of 45 AI ethics syllabuses made publicly available online were examined. Some important features captured from each syllabus were the course description, topics, department, and year. We observed overarching patterns across the AI ethics syllabi through supervised and unsupervised clustering and Latent Dirichlet Allocation (LDA) analysis. Some of these included information across various academic departments and the pre-post Chat-GPT era. This study is insightful as it offers a baseline for investigating various AI ethics topics that are described in academic departments, as well as uncovering potential gaps in the contents of AI ethics syllabi.

Introduction

The emergence of new technologies brings about increased responsibilities, such as the ethical implications of artificial intelligence (AI). Numerous professional organizations, such as IEEE, have begun to develop AI ethics guidelines. In addition, academia can be vital for fostering innovation and producing individuals skilled in both technical and ethical aspects [1]. Towards this goal, educators can prepare students to be culturally responsive, have a collaborative mindset with interdisciplinary skills, and be aware of equity issues and other social problems that plague society. A Stanford University Cyberlaw article [2] suggests several actions universities can take to support AI ethics: identifying key issues, promoting an ethical vision, and implementing ethical practices in teaching and research.

Examining AI ethics syllabi using natural language processing (NLP) clustering techniques can offer insights into the current state and areas for improvement in AI ethics education. This research sought to explore the contents of AI ethics syllabi to infer patterns and trends in higher education. By analyzing syllabus data, academic institutions can enhance their curricula to better address ethical considerations in AI development and deployment. Therefore, this research aims to answer the following question: *What can be inferred when exploring the contents of AI ethics syllabi using NLP clustering techniques?*

Methodology

Data Collection

Data was gathered through a strategic online search for publicly available syllabi and a few from the Tech Ethics Curriculum spreadsheet by [3]. Course description, topics, department, year, and public/private university were some of the important features that were collected and organized in a CSV file, where each row was a syllabus, and the columns were features and contents of the syllabus. Massive open online courses (MOOCs) were excluded, as well as syllabi that had missing data like department, university name, year, or location. A total of 45 syllabi were collected and included in this analysis.

Data Cleaning and Preprocessing

Regarding our important features, there were two cases where course descriptions were missing. In those cases, we used the course topics to replace the missing course descriptions. After the dataset was cleaned, we used the tweet-preprocessor package in Python, which allowed for tokenization, parsing, and removal of URLs and reserved words. The NLTK package was used to remove stop words. After those steps, the text was ready for embedding. We tested our data performance on sentence-transformer embeddings and term frequency-inverse document frequency (TF-IDF). We found that sentence transformers performed better with the clusters by having higher silhouette scores and cluster distances. We also compared the performance between course descriptions and topics. Topics showed greater similarities, and therefore, topics were prioritized to uncover uniqueness in syllabi in the analysis unless otherwise specified.

Results

K-means Clustering was used to find any patterns in terms of polarity among the syllabus topics. We tested values for k from 2 to 10 using the silhouette score, inter-cluster distances, intra-cluster distances, and cluster size as the evaluation metrics. We found that k = 2 was the best value for k, which was therefore used for this analysis.

As shown in Figure reffig:topicClusters, these 15 words are unique across the clusters. This means they do not reside in both clusters. Using k = 2, we found that the syllabi contents were separated into those that are more general or philosophical and those that are technical or specific ethical concepts.

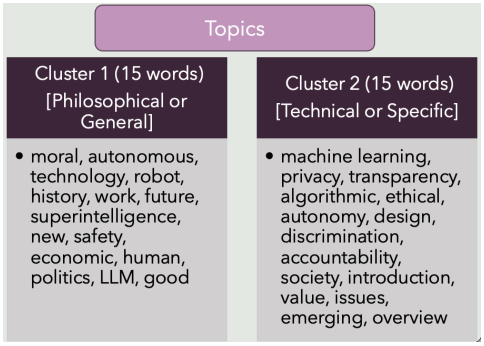

Figure 1: Topics of the two clusters

Supervised Clustering was conducted based on features such as department, year, and pre-post ChatGPT era. As shown in Figure 2, there is a topical vastness of AI ethics as well as how it is distributed across departments offering an AI ethics-related course. In addition, topics changed in the post-ChatGPT era (See Figure 3).

Figure 2: Topical range and distribution across departments offering an AI ethics-related course

Latent Dirichlet Allocation (LDA) Analysis was performed on k = 2, given that it was the best k in the K-means clustering. Python LDAvis library was used. The LDA analysis helped us see where the topics converged across all the syllabi (see Figure 4). We saw that *Data*, *Bias*, *Privacy*, and *Fairness* were the top converging topics across the syllabi, suggesting them as a priority for AI ethics.

Discussion and Implications

This approach utilized NLP techniques to analyze patterns across AI ethics syllabi, revealing that topics involving lists

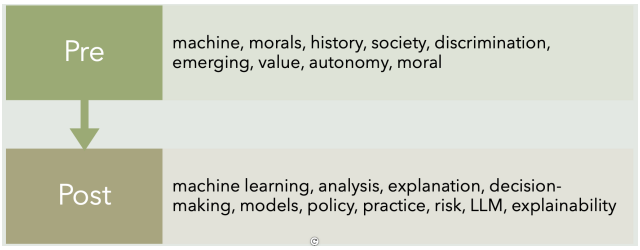

Figure 3: The topics covered pre-post Chat-GPT era, where the post-chat-GPT era is defined as 2023 and onwards

Cluster 1	Cluster 2
Data	Bias
Bias	Data
Fairness	Privacy
Privacy	Fairness
Transparency	Emerging

Figure 4: Topics across the syllabi

of words or phrases exhibited more distinctness than course descriptions, which may contain more noise. K-means clustering with k = 2 indicated polarity in concepts based on sentence-transformer embeddings. Supervised clustering revealed differences and patterns across departments and the pre-post Chat-GPT era. LDA analysis showed privacy, bias, fairness, and data as key issues across AI ethics syllabi, aligning with global literature on AI ethics principles [4]. Other researchers can replicate this methodology to analyze syllabi, aiding university professors and curriculum developers in identifying potential gaps and critical topics. The study's broader impact encompasses enhancing understanding of the sociotechnical system of AI ethics and the interdisciplinary nature of the field, as well as bridging gaps between academia and industry through student learning.

Limitations

K-means clustering requires specifying K, making it vulnerable to outliers and potentially skewing results. Analyzing clusters can be challenging due to this skew. Similarly, LDA analysis, while showing topic distribution across clusters, can be difficult to interpret.

Conclusion and Future Direction

The approach highlights the utility of NLP clustering analysis in this domain, revealing distinct patterns in AI ethics syllabi. K-means clustering demonstrates a polarization between general/philosophical and specific/technical topics. Supervised clustering uncovers additional patterns, suggesting implications for interdisciplinary collaboration and technological advancement. LDA topic modeling findings align with globally converging critical areas of AI ethics.

For future work, other clustering techniques, such as agglomerate clustering, can be applied to find more patterns in the dataset. More insights can be derived through using other NLP techniques, such as word similarities, to assess to what extent syllabi cover various topics.

References

[1] S. S. Bhullar, V. K. Nangia, and A. Batish, "The impact of academia-industry collaboration on core academic activities: Assessing the latent dimensions," *Technological Forecasting and Social Change*, vol. 145, pp. 1–11, 2019.

[2] B. W. Smith, "An Academic Vision for AI Ethics — cyberlaw.stanford.edu," https://cyberlaw.stanford.edu/blog/2023/04/academic-vision-ai-ethics, April 2023.

[3] C. Fiesler, "Tech ethics curricula: A collection of syllabi," Jul 2022. [Online]. Available: https://cfiesler.medium.com/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18

[4] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature machine intelligence*, vol. 1, no. 9, pp. 389–399, 2019.