Discovery of Research Trends in Computer Science Education on Ethics Using Topic Modeling

Sarah Parsons
Department of Computer Science
Wake Forest University
Winston-Salem, North Carolina, USA
sarah.parsons@alumni.wfu.edu

Natalia Khuri
Department of Computer Science
Wake Forest University
Winston-Salem, North Carolina, USA
natalia.khuri@wfu.edu

***Abstract*—To evaluate how ethics is addressed in computer science education, unsupervised text mining was performed to identify six salient topics and their trends in educational research publications. The majority of articles focused on the integration of ethics into introductory computer science courses, in particular those targeting non-majors. Models derived from a larger data set of research abstracts on ethics in computing, revealed additional topics in human-computer interaction and artificial intelligence, currently underrepresented in computer science education research. The agreement between automatic and manual topic assignments was 89%. Therefore, our approach provides a time-efficient and reproducible evaluation system to support curricular decision-making.**

***Index Terms*—ethics, computer science education, text mining, topic modeling**

I. Introduction

Technology has and will continue to serve as a source of progress and improvement in modern day society. As with all good things, it is important to consider the consequences or potential harm that may result from the adoption of new technologies. Rooted in the underpinnings of philosophy, ethics can be defined as a body of morals that guide the decisions and actions of an individual. Ethics pervades a broad range of topics and in the domain of computer science (CS), its meaning assumes a role of evolving responsibility. The Association for Computing Machinery (ACM) Code of Ethics provides a general framework for professional ethical conduct within the computing industry and describes basic moral principles for individuals and organizations. While these principles are intuitive in many situations, developers, product designers, and executive leaders face constant pressure as their morals are questioned when considering the ethical implications of their decisions.

As a secondary consideration in CS curricula, ongoing debate on whether or not ethics should be of higher priority has come to the forefront. Discussion on how to better incorporate ethics into CS education has led to new course offerings and higher expectations of recent graduates entering the technology industry [1]–[6]. A recent study of CS syllabi revealed that ethics of technologies is taught in various disciplines with over half of classes covering law and policy, privacy and surveillance, or philosophy [5].

Efforts to update the CS curriculum amidst the evolving atmosphere must involve a comprehensive survey of available instructional materials and best pedagogical practices. Bibliometric analysis is a commonly used technique of evaluating scholarly works. Evaluation results may be used for finding gaps and unmet needs within a discipline. In CS, recent bibliometric analyses studied relationships between digital citizenship and education [7], traced evolution of cyberbullying [8], examined historic and future representation of women in research [9], and evaluated long-term trends in CS education research [10]. For example, an analysis of articles on educational technology published in the last 50 years revealed that five most popular areas of educational research were technology-enhanced classroom pedagogy, project-based learning, blended learning, online social communities, and problem-based learning [11].

Retrospective surveys of educational publications provide insights into evidence-based practices and should be part of academic decision-making. However, bibliometric analyses are mostly done manually, which limits their application to relatively small data sets. Recent advances in text mining and the increased availability of research abstracts may allow for an automated and unbiased discovery of salient topics and trends within educational research.

Therefore, the main objective of our work was to use text mining for the discovery of ethics-related topics within two collections, namely, publication abstracts by the Special Interest Group on Computer Science Education (SIGCSE) and abstracts from the Semantic Scholar Open Research Corpus (S2ORC). In the analysis of results, we aimed to answer three specific research questions:

1) What are the major research topics evident in CS education research?
2) How has the trend of ethics in CS education evolved over time?
3) How does this trend compare and contrast with a broader view of ethics-related topics in computing?

We discovered main categories of ethics-related publications in both data sets and studied their evolution over time. Our results demonstrate that although overlap exists between topics derived from SIGCSE and S2ORC abstracts, some topics

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00166

were not well-represented in SIGCSE abstracts. Moreover, our results indicate that only a small minority of published educational papers evaluated pedagogical approaches to teaching ethics, and the majority of publications focus on the discussion of ethics in introductory courses, including those targeting non-majors. These findings may provide insights for future curricular and pedagogical initiatives.

The accuracy of our topic modeling was high, and the agreement between automated and manual topic assignments of SIGCSE abstracts was 89%. To the best of our knowledge this is a first application of text mining to ethics-related educational research. The proposed approach is time-efficient, reproducible and comprehensive. Our work contributes to the development of automated tools that augment human curricular decision-making.

The remainder of this paper is organized as follows. In Section II, we present the overall workflow of our modeling approach, followed by the details of the data sets and methods. Results are described in Section III and their broader context is discussed in Section IV. We conclude with a summary of our contributions in Section V.

II. Data and Methods

Topic modeling is an unsupervised text mining method used to infer topics from a corpus of text. These topics are identified based on patterns of words discovered within a subset of texts that embody their overall subject. Our modeling workflow (Figure 1) comprises four steps: (1) data collection and filtering; (2) text preprocessing; (3) vocabulary construction; and (4) topic modeling. In addition, results of SIGCSE topic modeling were manually reviewed and validated.

A. Data Collection and Preprocessing

We assembled two data sets comprising ethics-related abstracts from two libraries: the ACM Digital Library (DL) [12] and S2ORC [13]. DL is the major resource of publications in computing, containing journal articles, conference proceedings, and other scientific and research works. Within this library, SIGCSE proceedings offer a breadth of literature on CS education and they were selected as the first data set. Using the online DL interface, we set the filter to 'Research Article' content type. An additional filter was applied to identify research articles with the keyword 'ethics' in any of the search fields, such as title, abstract, full text, and so on. Abstracts and metadata of each article were then manually copied and saved in a comma-delimited file; these abstracts spanned the period from 2008 to 2020. Our search did not return any ethics-related SIGCSE publications prior to 2008.

The second data set comprised articles from S2ORC, encompassing publicly available scholarly works. S2ORC metadata, including abstracts are stored in a structured format, and the S2ORC is available for download in the form of 187 zip files. Metadata were first filtered with the 'fieldOfStudy' attribute set to 'Computer Science'. Next, retrieved files were parsed to extract publication year, title, abstract, and other fields. Notably, retrieved abstracts were drawn from many computing-related disciplines, including abstracts from several CS education journals and proceedings. Finally, we filtered the articles by the keyword 'ethics' in the title or abstract.

B. Vocabulary Construction

To prepare the data sets for modeling, all abstracts were preprocessed and tokenized. In preprocessing, we converted all words to lowercase, removed punctuation, numbers, and tokenized each abstract using white space characters as delimiters. Next, stop words were discarded using the predefined word list in the nltk library [14] and custom-made lists, consisting of words such as 'use', 'paper', 'abstract', 'ethics', and so on. We lemmatized all tokens using nltk's WordNetLemmatizer and stemmed them using nltk's SnowballStemmer. In this work, we excluded tokens with fewer than 3 characters. Abstracts with more than 20 tokens were used to create vocabularies for each data set. We filtered the vocabularies to remove noninformative tokens. Specifically, SIGCSE tokens that appeared in fewer than 10 documents or in more than 70% of documents were removed. S2ORC vocabulary comprised of tokens that appeared in at least 35 documents, but in no more than 70% of the documents. Finally, each tokenized abstract was represented by a bag-of-words (BOW) vector, comprising the counts of the occurrences of vocabulary tokens within each abstract.

C. Topic Modeling

In this work, two topic modeling approaches were used, namely Latent Dirichlet Allocation (LDA) and Dynamic Topic Modeling (DTM). All computer code was implemented in Python programming language, and executed on a High Performance Computing Cluster using one GPU. Approximately 200 megabytes of memory was required to preprocess and train the DTM model, while around 80 megabytes of memory was needed for the LDA model.

1) Latent Dirichlet Allocation: LDA is a popular hierarchical Bayesian approach that models a data set of texts as a random mixture over a finite underlying set of topics. Each topic is, in turn, characterized by the distribution of the vocabulary words [15]. Documents are then represented by the topic probabilities that comprise them.

We trained LDA models using the `sci-kit learn` library [16] and tuned three parameters: the number of topics, the document topic prior (alpha), and the word topic prior (beta). All other parameters were set to the default values of the `sci-kit learn` model. To find the best number of latent topics, we varied the number of topics from 5 to 10, and the values for alpha and beta parameters from 0.00001 to 0.001 and from 0.01 to 0.5, respectively.

2) Dynamic Topic Modeling: Originating from LDA, DTM identifies latent topics within a data set for a given set of time slices. As topics change in relevance and popularity over time, DTM uncovers the topic proportions in each time period. The time slices for SIGCSE ranged from 2008 to 2020 and included one year per slice. For S2ORC, the time slices ranged from 1970 to 2020, and they were appropriately sized to ensure there was a balanced distribution of publications in each time

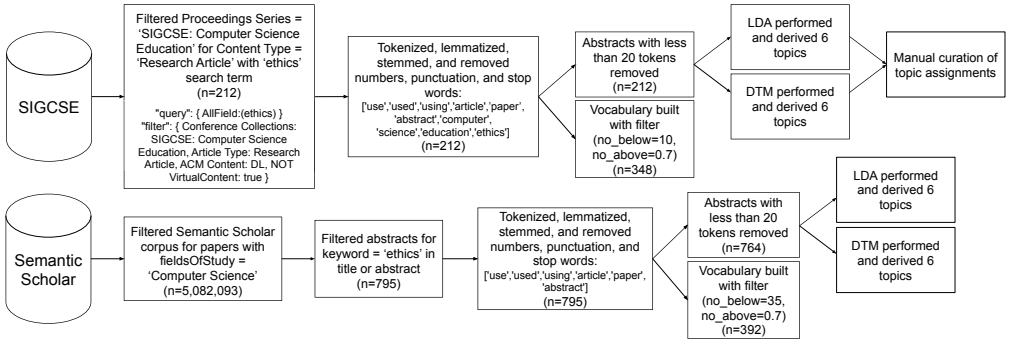

Fig. 1. Workflows for topic modeling of abstracts. Shown are the steps of the workflow comprising data collection, filtering, preprocessing, and topic discovery. Top. Modeling workflow of SIGCSE abstracts. Last step describes manual validation of NLP results. Bottom. Modeling workflow of S2ORC abstracts.

period. These time slices were divided as follows: [1970-1995, 1996-2000, 2001, 2002, ..., 2020]. To tune DTM parameters, we varied the number of topics from 5 to 10.

DTM was implemented using the `gensim` library in Python [17].

D. Manual Validation

We validated SIGCSE topic modeling results by manually examining all SIGCSE abstracts (n=212) and dividing them into six groups. Each group was manually titled and the documents from the manually curated groups were compared to the results of LDA and DTM. We calculated the agreement between the manual assignment of topics to the assignments of topics by the LDA and DTM models.

III. Results

Overall, our results demonstrate that topic modeling is a viable approach of discovering ethics-related topics in abstracts of research and scientific publications. There was a general agreement between LDA and DTM models about how research abstracts should be partitioned into topic categories. Additionally, automated topic assignment was compared with manual document annotations and was found to be consistent.

A. Topics and Trends in SIGCSE data set

Of the 2,062 SIGCSE research articles, spanning from 2008 to 2020, 212 were ethics-related. With approximately 6 sentences, on average, each tokenized abstract comprised an average of 86 words. A rising trend in ethics-related publications occurred in the last 3 years with twice the number of publications in 2020 (n=50) as in 2019 (n=29). A small spike in publications was also evidenced in 2010 (Figure 2).

Six main groups of publications were found by LDA and DTM models, and they were manually titled based on the most representative keywords within each group. Four of the six groups were identified by both, LDA and DTM, and they were software and teaching, social impact, security, and curriculum, while two were unique to each model. We examined unique document topics further. In the LDA model, such unique topics focused on 'data' and 'skills'. However, in the DTM model keywords used to describe these topics were bundled together into one topic, which we titled 'data and skills'. This topic represents research on data analysis, career preparation for students interested in industry, and CS skills, such as software development and programming. Finally, the DTM model identified an extra cluster of documents comprising abstracts on issues related to diversity and inclusion, which were not identified by LDA.

After manually reviewing each SIGCSE abstract and comparing document assignments with those assignments of LDA and DTM, we settled on the following six categories:

- software and course design (n=56)
- curriculum (n=55)
- social impact (n=36)
- professional development (n=27)
- data science (n=19)
- security (n=19)

Software and course design can be described as research on teaching, software applications, and on the incorporation of software and games in pedagogy. Social impact and diversity and inclusion represent the human component in CS, and the need for continuous advocacy for underrepresented groups in the field. Security and curriculum represent two important topics prevalent throughout the publication history of SIGCSE.

We examined trends of different SIGCSE topics. The trends varied for different topics (Figure 2). For example, the security topic was cyclical in nature through 2016, where there were no publications in some years, followed by years with a large number of publications. After 2016, the number of publications gradually declined before peaking again in 2019. Prior to 2015, research on preparing students with professional development was consistently discussed each year; however, in 2015 and 2016, no abstract on professional development was found. Following 2016, the number of publications on professional development remained the same. Further, there existed a lack of publications on data science in 2011 and 2012, as well as in the period from 2015 to 2017. There was a rapid growth in the number of data-related publications in 2020. Software, course design, and curriculum maintained a steady presence throughout the entire timeline. Likewise,

social impact was prevalent every year, with the exception of 2016. Moreover, a dramatic increase in research on data science, curriculum, and software and course design was seen in the last three years. In 2020, software and course design, data science, social impact, and curriculum topics spiked compared to previous years.

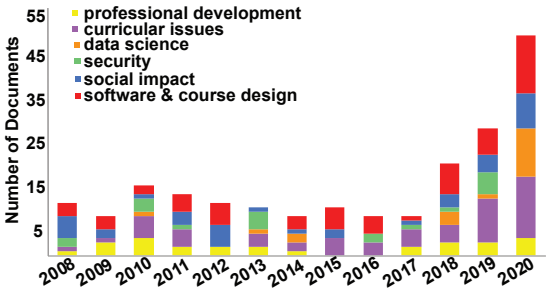

Fig. 2. Distribution of 6 topics by document count derived from LDA, DTM, and manual review for SIGCSE articles from 2008 to 2020.

B. Topics and Trends in S2ORC data set

Of the 5,082,093 CS abstracts in S2ORC, 795 were ethics-related. After preprocessing and filtering, the final set of documents was narrowed to 764 abstracts. Each abstract consisted of approximately 8 sentences and 104 words, on average.

There was a steady increase in ethics-related papers published in S2ORC from the twentieth century to the twenty-first century (Figure 3A). Furthermore, spikes in the number of publications occurred during 1990, 1994, 1998, and 2011, with the peak number of documents being published in 2016 (n=52).

Six topics were identified in S2ORC using LDA, each representing between 111 and 148 abstracts. Topic labels were selected based on the manual review of the top 50 keywords outputted by the model for each topic. The six S2ORC topics were:

- data and design (n=148)
- health and bioethics (n=146)
- engineering and education (n=133)
- artificial intelligence (AI) and human-computer interaction (HCI) (n=114)
- social issues (n=111)
- professional code of ethics (n=111)

Although LDA topic modeling produced a static view of the S2ORC, trends were identified by computing the proportion of documents in each topic in different time periods. Notably, the number of abstracts in 'data and design' aligned with the increase in overall publications. A similar trend was observed for health and bioethics. Two other topics, professional code of ethics and AI/HCI, increased but on a smaller scale (Figure 3A).

To evaluate this further, we used the DTM algorithm to derive salient topics in the S2ORC data set, which incorporates time into topic modeling. We balanced the distribution of documents across 22 time slices. Similar to LDA, DTM also identified six topics within S2ORC. We examined the keywords defining each topic and compared them to the keywords of LDA topics. Although there was an overlap between the outputs of the two algorithms (AI/HCI, engineering and education, and professional code of ethics), some topics were unique to each model (Figure 3B).

IV. Discussion

Systematic bibliographic reviews of research and scientific publications are conducted regularly [10], [11], [18]. Such reviews may uncover latent topics in large corpora and their trends, and they may find gaps and unmet research needs. Additionally, such reviews may provide insights for academic decision-making on minor and major curricular updates, changes to course content and the adoption of evidence-based pedagogical practices. Advances in text mining technologies can automate such discoveries and streamline human practices. In this work, we used text mining to analyze ethics-related scholarly abstracts and made three research contributions.

First, we determined that topic modeling algorithms can accurately cluster publication abstracts into groups of closely related topics. We focused on the discovery of topics in ethics-related abstracts published in CS conference proceedings and journals. We found that DTM models were more accurate than LDA. Moreover, DTM topic assignments were in high agreement with manual topic assignments of SIGCSE abstracts (89%). Given the rapid growth of new publications and the digitization of older publications, our results indicate that such algorithms can be successful in the analysis of research publications.

Second, we discovered that abstracts of SIGCSE publications that focused on ethics, can be divided into six groups, namely, data science, professional development, software and course design, curricular issues, social impact, and security (Figure 2). Overall, our results indicate that the number of SIGCSE publications on ethical issues is increasing. However, the number of S2ORC publications on ethics peaked in 2016, and it remained stable. However, only 13 of the 212 SIGCSE abstracts describe pedagogy of teaching ethics.

For SIGCSE, 2018 saw the largest rise in publications and served as the onset of a growing trend in CS ethics research. Topics covered in 2018 included cybersecurity courses [19], [20], service learning [21], [22], methods for introducing the 'new' subject of data science [23]–[25], and a strong focus on improving the experience of underrepresented students, such as non-majors, associate-degree transfers, and students with disabilities [26]–[29]. In the last three years, a large focus was on data, software and course design (often focused on programming with Python) [30]–[32], and ethics curriculum improvements in response to the update of the ACM Code of Ethics [33], [34]. Although continuing to grow and achieving peak interest in 2020, research on social impact and professional development was not as prevalent as other topics. Also, after peaking in 2019, security was not discussed in 2020.

Intrigued by the trends discovered using dynamic topic modeling of S2ORC abstracts, we manually examined them.

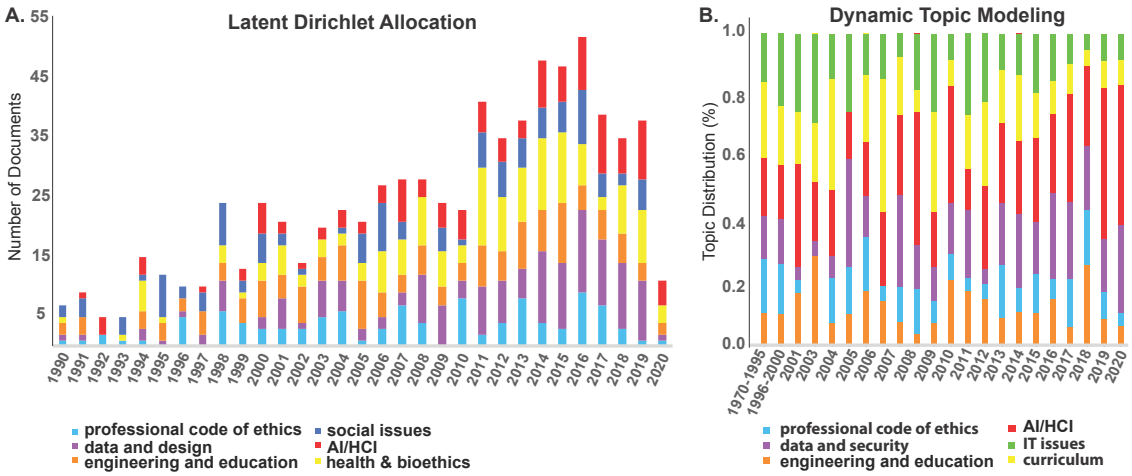

Fig. 3. Distribution of S2ORC topics and trends modeled using LDA and DTM. Timeline is shown on the x-axis and topics are denoted by colors. A. Number of documents in each topics discovered using LDA. B. Distribution of topics identified using DTM.

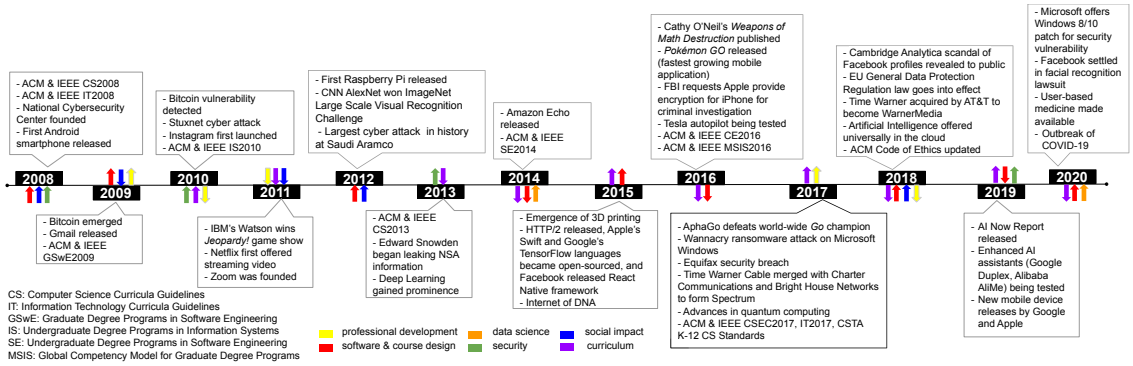

Fig. 4. Timeline of SIGCSE topic trends mapped to real-world events with indicators of primary topics and their increasing and decreasing topic proportion based on prior year.

In 1990, a panel session was held to discuss the integration of ethics into computer science curriculum [35], while a paper was published regarding recent 'scams' and 'gates' emphasizing the need for a definition of ethics [36]. In S2ORC, ethics, at the time, was tied closely with business and the professional industry. Educational publications from those years called for standalone ethics courses in computer science curriculum with a recommendation that they be taught by philosophers or social scientists as opposed to computer scientists [37], [38]. Of particular interest in 1998, was the need for the analysis of ethics codes regarding common violations, such as copying software, hacking, virus spreading, sabotage, computer fraud, and privacy concerns with surveillance [39]–[41]. In 2011, publications began to focus on ethics in multidisciplinary computing fields, such as bioethics, IT ethics, and roboethics [42]–[44]. Similarly, SIGCSE publications in 2010 consisted of on curricular issues for interdisciplinary domains [45].

S2ORC represents a large body of information on computing, including applications in other domains. Therefore, some topics discovered in S2ORC, which were not represented by SIGCSE proceedings, serve as possible additions to CS curricula or courses. As Information Technology (IT) continues to branch into other domains, such as bioethics and medicine, incorporating this topic into CS curricula may be worth considering [46], [47]. For example, S2ORC corpus contained large clusters of publications about ethics in healthcare and life sciences (n=146), as well as about issues in AI and HCI (n=114). SIGCSE abstracts lacked these topics, calling for more attention to these issues within CS curriculum. Our analysis also supports the recent findings that CS courses on ethics, offered in the last few years, tend to focus on law and policy, or privacy and surveillance, but not as much on ethics of interdisciplinary topics [6].

Additionally, S2ORC publications on professional code of ethics (n=111) were clustered in a well-defined group (Figure 3). However, publications on professional code of ethics found in SIGCSE did not cluster together in our analyses.

This may call for greater coverage of the code of conduct within the CS curriculum, and for more research on best practices for teaching it. Similarly, while SIGCSE publications discuss professional development and industry preparation for students, the S2ORC topic of professional code of ethics extends this discussion and incorporates the importance of adherence to a code of ethics in IT organizations and other computing businesses.

Finally, the growing trend of SIGCSE publications focused on CS curriculum, software, and course design, did not fully align between the two data sets. SIGCSE publications on curricular issues increased steadily from 2016 to 2020, whereas S2ORC yielded consistent output in the twentieth century, then a cyclical trend before gradually declining after 2012.

To examine whether there were specific real-world event triggers of the trends in ethics-related SIGCSE publications, we developed a timeline of important events (Figure 4). Security vulnerabilities and the cyclical nature of security research appear to coincide, based on events occurring prior to the increase in interest in security. For example, the 2013 increase in security research followed the largest cyber attack in history, occurring in 2012. Likewise, software research tended to increase during the years of new software releases, such as in 2008, 2012, and 2015. Finally, a peak of abstracts devoted to ethics in 2020, closely followed the update of the ACM Code of Ethics in 2018.

In summary, there are three take-away points of our analyses. First, the number of publications on ethical considerations remains very low in computer science education research as well as in computing research, in general. As expected, education research articles are frequently inspired by real-world events or major curricular updates, and there is a publication lag between such events and the dissemination of research results. S2ORC publications are more in-step with the current events, as evidenced with a large representation of abstracts discussing ethics of AI and HCI, as well as diversity and inclusion. Third, most SIGCSE publications include ethical issues as a desirable outcome of some other pedagogical intervention, especially in the introductory computer science courses.

There are some limitations to this work. First, we built models using only 212 SIGCSE and 764 S2ORC abstracts of research publications and their texts were very short (fewer than 250 words). However, DTM and LDA perform better with increasing amounts of text, and DTM requires a balanced data set across time slices to produce accurate results. Therefore, when full texts become available, they will be used to improve our models.

V. Conclusion

In Computer Science, ethics has been under discussion for several decades, and it continues to be integral to the academic and industry communities. In this research, automated topic modeling was used on two data sets of CS publications, namely, SIGCSE, focused on education, and S2ORC, covering a broad range of CS domains. By successfully employing topic modeling to discover topics and trends of ethics in CS education over the past five decades, insights were gained into the various concentrations of ethics and the periods during which these patterns shifted. The results of our research indicate that despite a clear need for incorporating ethics across the entire CS curriculum and for new pedagogical methods, the majority of educational research focuses on introductory courses, data science and computer security. Thus, there exist opportunities to introduce into curricula, additional topics, such as professional code of ethics, healthcare and life sciences, as well as AI, HCI, and IT ethical issues.

While focused on research on ethics education, our proposed approach can be used to automatically evaluate large collections of research abstracts to discover salient topics and their trends. The bibliographic evaluation system proposed in this work is reproducible, time-efficient and unbiased, and it can be used to augment human decision-making on curricular issues.

Acknowledgment

This publication resulted (in part) from research supported by the Seed Award of the National Center for Women & Information Technology (NCWIT). The authors thank Tian Yun and Cody Stevens for the assistance with the execution of computational experiments, and acknowledge the Distributed Environment for Academic Computing (DEAC) at Wake Forest University for providing HPC resources that have contributed to the research results reported within this paper.

References

[1] F. Grodzinsky, E. Gehringer, L. S. King, and H. Tavani, "Responding to the Challenges of Teaching Computer Ethics," *SIGCSE Bull.*, vol. 36, no. 1, pp. 280–281, Mar. 2004.

[2] C. Spradling, L.-K. Soh, and C. Ansorge, "Ethics Training and Decision-making: Do Computer Science Programs Need Help?" *SIGCSE Bull.*, vol. 40, no. 1, pp. 153–157, Mar. 2008.

[3] D. Gotterbarn, "Thinking Professionally: Professional Computer Ethics: "I Didn't Do It" is Not Good Enough," *SIGCSE Bull.*, vol. 41, no. 2, pp. 65–66, Jun. 2009.

[4] D. K. Larson and K. W. Miller, "Action Ethics for a Software Development Class," *ACM Inroads*, vol. 8, no. 1, pp. 38–42, Feb. 2017.

[5] B. J. Grosz, D. G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons *et al.*, "Embedded EthiCS: Integrating Ethics across CS Education," *Commun. ACM*, vol. 62, no. 8, p. 5461, Jul. 2019.

[6] C. Fiesler, N. Garrett, and N. Beard, "What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, ser. SIGCSE 20. New York, NY, USA: Association for Computing Machinery, 2020, p. 289295.

[7] J. S. Fernández-Prados, A. Lozano-Díaz, and C. Cuenca-Piqueras, "Digital Citizenship and Education: State of the Art and Bibliometric Analysis," ser. ICIET 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 174178.

[8] M. J. González-Moreno, C. Cuenca-Piqueras, and J. S. Fernández-Prados, "Cyberbullying and Education: State of the Art and Bibliometric Analysis," ser. ICIET 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 191195.

[9] D. McKay and G. Buchanan, "Shaking the Tree: Understanding Historic and Future Representation of Women at OzCHI," in *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, ser. OZCHI'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 412417.

[10] Simon, "Twenty-Two Years of ACE," in *Proceedings of the Twenty-Second Australasian Computing Education Conference*, ser. ACE'20. New York, NY, USA: Association for Computing Machinery, 2020, p. 203210.

[11] X. Chen, D. Zou, and H. Xie, "Fifty years of British Journal of Educational Technology: A topic modeling based bibliometric perspective," *BJET*, vol. 51, pp. 692–708, 2020.

[12] "ACM Digital Library," 2020. [Online]. Available: https://dl.acm.org/

[13] A. Waleed, D. Groeneveld, C. Bhagavatula, I. Beltagy, M. Crawford, D. Downey *et al.*, "Construction of the Literature Graph in Semantic Scholar," in *NAACL*, 2018.

[14] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, 1st ed. O'Reilly Media, Inc., 2009.

[15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, no. null, p. 9931022, Mar. 2003.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, May 2010, pp. 45–50, http://is.muni.cz/publication/884893/en.

[18] X. Chen, D. Zou, G. J. Cheng, and H. Xie, "Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education," *Comput. Educ.*, vol. 151, p. 103855, 2020.

[19] M. R. Asghar and A. Luxton-Reilly, "Teaching Cyber Security Using Competitive Software Obfuscation and Reverse Engineering Activities," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 179–184.

[20] G. Jin, M. Tu, T.-H. Kim, J. Heffron, and J. White, "Game based Cybersecurity Training for High School Students," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 68–73.

[21] L. Pollock, J. Atlas, T. Bell, and T. Henderson, "A Computer Science Study Abroad with Service Learning: Design and Reflections," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 485–490.

[22] S. Robinson and M. Hall, "Combining Agile Software Development and Service-learning: A Case Study in Experiential IS Education," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 491–496.

[23] J. S. Saltz, N. I. Dewar, and R. Heckman, "Key Concepts for a Data Science Ethics Curriculum," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 952–957.

[24] A. Dryer, N. Walia, and A. Chattopadhyay, "A Middle-School Module for Introducing Data-Mining, Big-Data, Ethics and Privacy Using RapidMiner and a Hollywood Theme," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 753–758.

[25] S. Dahlby Albright, T. H. Klinge, and S. A. Rebelsky, "A Functional Approach to Data Science in CS1," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 1035–1040.

[26] S. Narayanan, K. Cunningham, S. Arteaga, W. J. Welch, L. Maxwell, Z. Chawinga *et al.*, "Upward Mobility for Underrepresented Students: A Model for a Cohort-Based Bachelor's Degree in Computer Science," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 705–710.

[27] J. Q. Dawson, M. Allen, A. Campbell, and A. Valair, "Designing an Introductory Programming Course to Improve Non-Majors' Experiences," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 26–31.

[28] C. Tang, C. Tucker, C. Servin, and M. Geissler, "Computer Science Curricular Guidance for Associate-Degree Transfer Programs," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 435–440.

[29] S. Ludi, M. Huenerfauth, V. Hanson, N. Rajendra Palan, and P. Garcia, "Teaching Inclusive Thinking to Undergraduate Students in Computing Programs," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '18. Baltimore, Maryland, USA: Association for Computing Machinery, Feb. 2018, pp. 717–722.

[30] A. C. Bart, A. Sarver, M. Friend, and L. Cox II, "PythonSneks: An Open-Source, Instructionally-Designed Introductory Curriculum with Action-Design Research," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '19. Minneapolis, MN, USA: Association for Computing Machinery, Feb. 2019, pp. 307–313.

[31] B. Krupp and A. Watkins, "CS0: Introducing Computing with Raspberry Pis," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '19. Minneapolis, MN, USA: Association for Computing Machinery, Feb. 2019, pp. 832–838.

[32] T. Babbitt, C. Schooler, and K. King, "Punch Cards to Python: A Case Study of a CS0 Core Course," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '19. Minneapolis, MN, USA: Association for Computing Machinery, Feb. 2019, pp. 811–817.

[33] M. J. Wolf, D. Gotterbarn, and M. S. Kirkpatrick, "ACM Code of Ethics: Looking Back and Forging Ahead," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '19. Minneapolis, MN, USA: Association for Computing Machinery, Feb. 2019, pp. 801–802.

[34] R. Blumenthal and J. Blumenthal, "Consider Visualizing Society within the ACM Code of Ethics," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1292.

[35] D. Gotterbarn, D. G. Johnson, K. W. Miller, and E. H. Spafford, "Methods of integrating the study of ethics into the computer science curriculum (panel session)," in *SIGCSE '90*, 1990.

[36] N. Nagarajan, "What's computer ethics, anyway?" *ACM Sigsac Review*, vol. 8, pp. 24–32, 1990.

[37] S. C. Pulliam, "A social implications of computing course which teaches computer ethics," in *ECA '94*, 1994.

[38] D. G. Johnson, "Who should teach computer ethics and computers and society?" *SIGCAS Comput. Soc.*, vol. 24, pp. 6–13, 1994.

[39] G. T. Marx, "An ethics for the new surveillance (abstract)," in *ACM Policy*, 1998.

[40] J. Leiwo and S. Heikkuri, "An analysis of ethics as foundation of information security in distributed systems," *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, vol. 6, pp. 213–222, 1998.

[41] M. Woodbury, "Defining web ethics," *Science and Engineering Ethics*, vol. 4, pp. 203–212, 1998.

[42] N. de Vecchio Puplaksis, "Bioethics, knowledge construction and e-learning: new challenges to the Bioethics teaching (Bioética, construão de conhecimento e e-learning: novos desafios ao ensino da Bioética)," 2011.

[43] R. Homkes and R. A. Strikwerda, "Applying CPR to the teaching of IT ethics," in *SIGITE '11*, 2011.

[44] S. Bringsjord, J. Taylor, B. van Heuveln, K. Arkoudas, M. Clark, and R. L. Wojtowicz, "Piagetian Roboethics via Category Theory Moving beyond Mere Formal Operations to Engineer Robots Whose Decisions Are Guaranteed to be Ethically Correct," 2011.

[45] M. D. LeBlanc, T. Armstrong, and M. B. Gousie, "Connecting across campus," in *Proceedings of the 41st ACM technical symposium on Computer science education*, ser. SIGCSE '10. Milwaukee, Wisconsin, USA: Association for Computing Machinery, Mar. 2010, pp. 52–56.

[46] J. Anderson, "The role of ethics in information technology decisions: a case-based approach to biomedical informatics education," *International journal of medical informatics*, vol. 73, no. 2, pp. 145–50, 2004.

[47] C. Coelho and G. D. Kornis, "Toward a new frontier on economic and technologic assessment in intensive care: the role of ethics," *Critical Care*, vol. 7, p. 89, 2003.