Content Moderation and Personalization of Learning Materials with Learnersourcing

Ping Wang[1], Tianyi Li[2,*,†]

[1]Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030

[2]Purdue University, 401 Grant St, West Lafayette, IN 47907

Abstract

Implicit bias is commonly included unintentionally in the course materials from different aspects, such as gender, culture, ability, and occupation. Such biases can create barriers for engaging and retaining novice learners in the computing subjects. To address this challenge, current approaches have largely focused on evaluating and improving the quality of learning materials, which heavily relies on the expertise of instructors, researchers, and educators. In this position paper, we argue that students can also contribute to content moderation and further, guide the personalization of learning materials, using a learnersourcing approach. We envision the proposed approach can help address three substantive challenges: (1) isolated learning of students, (2) one-size-fits-all materials, and (3) possible implicit bias in learning materials.

Keywords

Content moderation, Learnersourcing, Implicit bias, Learning materials, Classification task

1. Introduction

Existing materials used in computer science classes often reinforce harmful stereotypes [1]. For example, Medel and Pournaghshband's work discussed three main types of gender inequality manifested in the learning materials, including (1) *representation*, that female names are disproportionately associated with negative roles in the teaching examples used in cryptography learning; (2) *imagery*, that male-dominant learning examples such as the image of Lena, objectify and project stereotypes against women; (3) *language* used in the learning materials, such as pronouns, usually carries negative connotation against women. In this work, we explore using learnersourcing to detect biases in learning materials from student-centered perspectives.

Learnersourcing is a pedagogically meaningful form of crowdsourcing where learners collectively contribute novel content while engaging in meaningful learning experiences themselves [2]. Peer assessment can be considered as a special instance of learnersourcing and has decades of history [3]. Recent research has successfully sourced high-quality multiple-choice questions [4], and programming assignments [5] and participated in the high-level planning and

The first annual workshop on Partnerships for Cocreating Educational Content, March 13, 2023, Arlington, TX

*Corresponding author.

†These authors contributed equally.

ping.wang@stevens.edu (P. Wang); li4251@purdue.edu (T. Li)

https://leafnlp.org/ping (P. Wang); https://tl-tianyi-li.github.io/ (T. Li)

0000-0002-0379-9183 (P. Wang); 0000-0002-1145-2526 (T. Li)

CEUR Workshop Proceedings (CEUR-WS.org)

organization of topics covered in a course [6], to name a few. In this position paper, we propose to learnersource the *content moderation* with students to identify biases against different student groups and backgrounds in learning materials by formulating it as a classification scenario in machine learning. Students are encouraged to challenge the existing knowledge-power structure and advocate their cultural and social identities, co-constructing a more democratic and inclusive learning community.

Content moderation has been a challenging and controversial practice as it both protects and constrains the community [7]. When applied to educational materials, which carry inherent power and authority in traditional classrooms [8], content moderation requires even more careful planning and consideration to ensure responsible and effective bias detection and mitigation outcomes. Learnersourcing allows student-centered content moderation against biases and triggering content for different cultural and social backgrounds, but also faces challenges and potential hazards. As is pointed out by Darvishi et al. [9], the quality of learnersourced content is usually varied and can be "ineffective, inappropriate, or incorrect". On the one hand, peer assessment alone can not reliably judge the quality of student-created materials due to the students' limited expertise, experience, and motivation to conduct such evaluation tasks [9]. On the other hand, students might have varied sensitivity to the implicit biases in the learning materials thus personalization will be crucial [10]. Glassman et al. demosntrated how learnersourcing can effectively support personalization of the hints needed by learners of different levels of experience [11]. Building on prior research, we lay out two initial considerations of the challenges in below, and hope to continue the discussion with the research community.

First, awareness and adequacy for assessing the cultural and social biases in learning content require long-term cultivation and development of multiple stakeholders (instructors, advisors, administrators, students, etc.) [12]. Brief instructions and tutorials included as part of learnersourcing tasks or workflows could be effective but are also highly customized and hard to generalize to other tasks or courses. Training on diversity, inclusivity and other relevant literacy is thus not optimal and not supposed to be considered in silos within one learnersourcing task or course, but rather should be involved as part of a curriculum or program level objective [13]. Unfortunately, such training is usually lacking in most STEM education programs or students do not have the incentives to engage in the available resources [14].

Second, there should not be any one-size-fits-all rules and standards for assessing biases in learning content [15, 16]. While distributing biased learnersourced content risks reinforcing unwanted stereotypes and prejudices, intervening content curation with one unified rule determined by one or a few people carries the same or even more risks. As a result, existing evaluation mechanisms (peer assessment, instructor-driven assessment, or automated assessments) all fall short of this unique personalization need. Specifically, peer assessment may not carry as many learning benefits in the context of content moderation and more importantly, may incur extra hazards due to conflicts of values and standards [17].

Addressing the first challenge requires institutional and long-term coordination and collaboration at different scales. The second challenge calls for learner-centered design and has great potential for human-AI teaming. This paper proposes a bottom-up approach that returns to the learnersourcing philosophy: by engaging students in personalized content moderation as a standalone learning task (second challenge). Students are engaged in some lightweight ethical training that can onboard and hopefully intrigues them to participate in future diversity and inclusivity training and endeavors. In this context, we highlight the importance and challenges of using learnersourcing to assess and critique the cultural and social biases of learning materials and discuss the following questions:

1. **Benefits:** What are the potential learning benefits of content moderation tasks?
2. **Context:** How to incorporate the learning benefits of content moderation into specific subjects and learning objectives in the context of different courses?
3. **Artifacts:** How to implement content moderation-supported learning tasks in course designs in a learnersourcing manner?
4. **Workflow**: How can learnersourced content moderation be included in existing learnersourcing workflows?

The rest of the paper is structured as follows. In Section 2, we describe a case example where we propose to use the learnersourcing approach to enable student-centered content moderation and teach classification algorithms. We first review the current practices and challenges in machine learning courses and propose a design of learnersourcing artifacts to facilitate the learning of classification algorithms with personalized content moderation. In Section 3, we suggest future research directions and our expected implications and broader impact of the proposed approach.

2. Learning Task Design: A Case Example with Classification Algorithms in Machine Learning

We propose one example where content moderation of learning materials can carry learning benefits. That is, in a machine learning class, learning classification algorithms by classifying (i.e., identifying) the possible biases in the learning materials. Prior research has revealed how different cultural biases against different genders, races, and other aspects are prevalent in online assessments [15], learning tools and software [16, 18], textbooks [19], visual content [20, 21] and curriculum design [22]. We will use such pre-existing learning materials from publicly available educational resources for students to classify. Using binary classification as an example, students can classify given learning materials into two categories: unbiased or biased, based on their own standards and perceptions of the learning materials. When designing the classification criteria, students will have the agency to focus on biases against their self-identified genders, races, and other cultural backgrounds. Students can then implement personalized classification algorithms to detect different types of biases.

2.1. Current Education of Classification Algorithms

In the big data era, machine learning is one of the most in-demand courses in computer science. Classification algorithms serve as one of the primary topics in machine learning-related courses where the goal is to analyze and categorize the given data into a class or category. Therefore, designing suitable teaching strategies for classification algorithms plays an important role in machine learning education. Here, we summarize two important strategies for classification

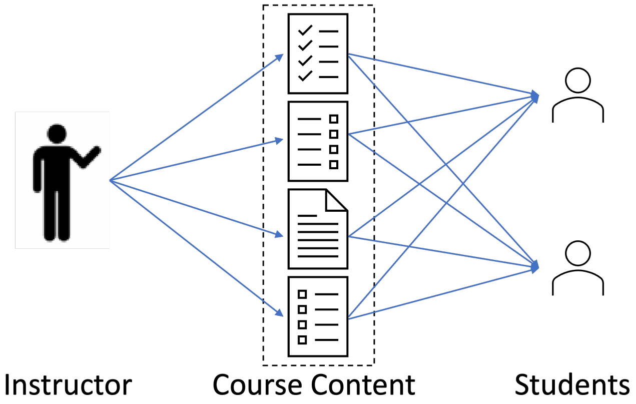

Figure 1: An illustration of current education practices. The instructor collects and constructs all course materials for all students in general. Therefore, there are still several challenges that need us to rethink, such as isolated learning of students, one-size-fits-all materials for all students, and possible implicit bias in educational materials. This motivates us to leverage the learnersourcing to perform content moderation as discussed in Section 2.2 and Figure 3.

algorithms in current educational practice (illustrated in Figure 1), which show great success in teaching classification algorithms.

S1. Preparing students with mathematical backgrounds. The knowledge of mathematics (e.g., calculus, linear algebra, and probability) is fundamental for understanding many concepts and methods in machine learning. Therefore, to help students get a deeper understanding of mathematical theories, instructors usually provide a review of the basic mathematic knowledge to prepare students with the necessary skills for advanced topics.

S2. Combining theory with practice. The problem formulation, optimization, and evaluation of classification algorithms usually include various notations, equations, and proofs, which are important tools to help students understand the theories behind each classification algorithm. However, our education practices show that these are exactly the most challenging part for students to understand and digest. Therefore, instructors usually include illustrative examples in lectures and programming questions about real-world problems in homework assignments to provide students with hands-on experiences and further enhance their learning of the theories for different classification algorithms.

However, to achieve better education outcomes and help more students to successfully learn classification algorithms, we argue that there are still several challenges in the current education practices that need us to rethink and further improve the current pedagogical strategies.

C1. Isolated learning of students. In current education strategies, even though there are student discussions both in-class and after class via online platforms, these discussions primarily focus on addressing clarification questions and heavily rely on the instructors to facilitate and and provide answers. Therefore, student learning is mostly performed in isolation and lacks peer support and a sense of community. We argue that collecting various opinions and feedbacks from students about the learning materials will not only enhance their understanding of the topics but also further improve the quality of the materials and benefit future students with learnersourced course materials.

C2. One-size-fits-all materials for all students. Typically, machine learning course materials for classification algorithms are designed to target the general student audience. However, we find that there are certain drawbacks to the one-size-fits-all course materials. First, the default learning examples may not be familiar or relatable to certain student populations and thus may limit their understanding of the classification algorithms. Second, such unequal familiarity and relatability may lead to unequal learning outcomes. For example, students who lack mathematic fundamentals and a machine learning background may find it difficult to follow complex problem formulations and optimizations, while students with strong backgrounds may be interested in learning more theoretical details. Therefore, we argue that personalized education of classification algorithms is essential for teaching machine learning. Providing course materials based on the background of individual students will help students learn more effectively.

C3. Possible implicit bias in educational materials. Implicit bias may be unintentionally included in the course materials from different aspects, such as culture, gender, ability, occupation, social status, and language, which may impact students' understanding of algorithms. For example, students who do not speak French will find it difficult to understand an example presented in French, and it may take more effort for students without any work experience to understand an example of classification about employee retention. All these implicit biases may affect students' learning experiences and outcomes. This motivates us to introduce learnersourcing to facilitate the content moderation of machine learning teaching materials to consider information about students' perception of biases in their learning materials and environment, their personal backgrounds and interests, as well as their concerns and needs during learning.

2.2. Moving Forward with Learnersourced Content Moderation

Motivated by the aforementioned challenges in the education of machine learning, we propose to perform the learnersourcing content moderation of course materials with deliberated learning task design that is integrated into the general machine learning pipeline (Figure 2). Specifically, we propose to include four primary steps, including collecting pre-existing learning materials, learnersourcing annotation of course materials, learning personalized classifiers, and identifying personalized course materials. These four steps are corresponding to the four important steps in general machine learning pipelines, including data collection, data annotation, training of algorithms, and application of the trained models, respectively. Figure 2 shows the overall framework of the proposed strategy and Figure 3 illustrates how the learning task contrasts the traditional practices in Figure 1. Therefore, the proposed learning task design not only provides a real application scenario for individual students to train the personalized classifiers but also enables students to identify personalized learning materials. We describe the details of each step in the designed learning task as follows.

Step 1. Collecting learning materials. In this study, we refer to learning materials as *self-contained information pieces* such as illustrative examples, tables, figures, and narrative texts. Based on this definition, we will first collect a set of learning materials from multiple resources, such as textbooks, online resources, and publicly available lectures. Our goal in this step is to collect a set of learning materials with diverse data types and learning contents for students to provide their feedback during the annotation step and further utilize for training classification

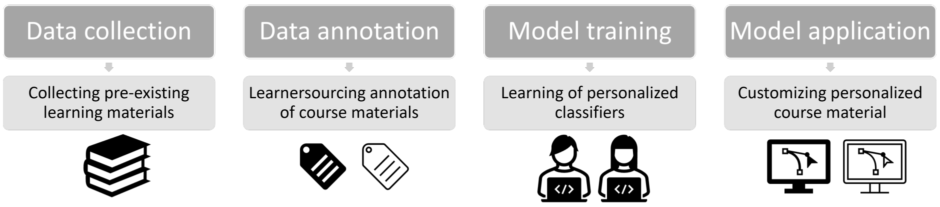
Data collection
Data annotation
Model training
Model application
Collecting pre-existing learning materials
Learnersourcing annotation of course materials
Learning of personalized classifiers
Customizing personalized course material

Figure 2: The overall framework of the proposed learnersourcing content moderation strategy, corresponds to the machine learning pipeline. Students are actively involved in every step. The instructor works with students collaboratively to collect various course materials. After that, students will label the learning content based on his/her own bias criteria. Each student then trains a personalized classifier which can be used to personalize course materials for themselves in the learning of other topics or other courses. More details of the proposed content moderation strategy are described in Section 2.2.

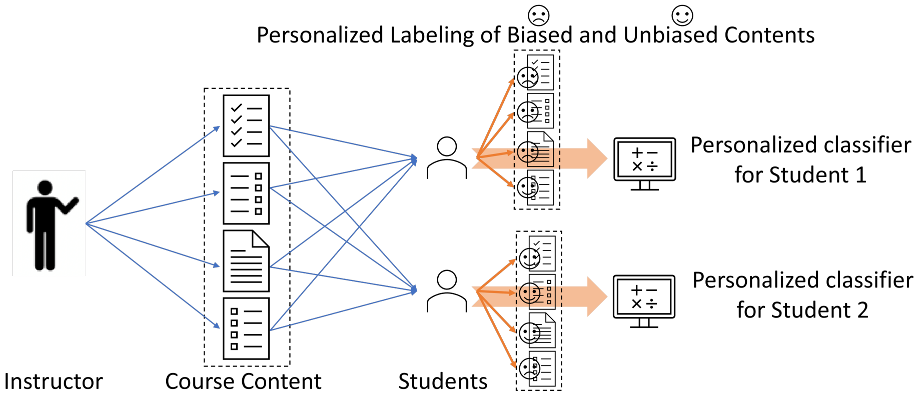
Personalized Labeling of Biased and Unbiased Contents
Personalized classifier for Student 1
Personalized classifier for Student 2
Instructor
Course Content
Students

Figure 3: Different from teaching all course materials to each student in traditional teaching shown in Figure 1, each student actively selects and annotates unbiased and personalized course materials and trains a personalized classifier based on different bias criteria.

models. Students will also be involved in this step to help them get a deep understanding of the data collection step.

Step 2. Learnersourcing annotation of course materials. This step aims to collect a binary label (unbiased or biased) for each course material collected in Step 1. Different students may have different opinions on the same course material based on their individual identities, prior knowledge, and social-economic status. Therefore, instead of obtaining a fixed class label for each course material, we propose to utilize personalized class labels for each student. To better track the specific criteria for each course material, we plan to conduct a survey to collect the commonly concerned aspects across different classes of students and consider these aspects as criteria candidates for students to indicate during annotation. The annotation criteria used by students will serve as an important feature when training the personalized classifiers. From the machine learning perspective, this step provides students with an opportunity to participate in obtaining the class label of the data, which will further reinforce their understanding of both the data input for classification tasks and the objective of classification in general. Besides, in this step, students can leverage the annotated data to practice data pre-processing and analysis, such as feature selection and extraction, feature normalization, and data visualization.

Step 3. Learning of personalized classifiers. The personalized annotated labels of the course materials in Step 2 can support each individual student to train the personalized classifier. Students can practice how to obtain important features for different data types and feed them as input to classification models. Students can investigate different machine learning models, the same model but with different parameter settings, or feed different features as input. Therefore, it provides more flexibility for students to explore different aspects of classification models, get a sense of evaluating their advantage and disadvantage, and further determine the best classifier for them to identify suitable course materials. Moreover, students also have the opportunity to explore machine learning models that can handle data in multiple modalities, such as tabular, texts, and images [23]. We expect that the learned classifier has the ability to capture the personalized underlying criteria and interests of each individual student.

Step 4. Identifying personalized course materials. The selected classifiers in Step 3 can serve as the personalized filter for each student to select suitable course materials. For any new course materials, students can feed them into their personalized filter to automatically check, which will be helpful for students to efficiently identify learning materials through the large volume of the material pool. Moreover, students can perform the filtering step without depending on other students and instructors, which will improve both the quality of the teaching materials in general and the efficiency of the teaching.

The deliberated design of the learning tasks helps students to go through every important aspect of the whole classification pipeline, including data collection, annotation, model training, and the application of the trained classification models. Different from passive learning in traditional education, the proposed learnersourcing content moderation strategy provides students with a real scenario strongly related to them to actively learn the classification task with personalized classifiers and further identify personalized learning materials. Moreover, the proposed strategy for content moderation will help the instructors of machine learning be aware of potential aspects that students consider during learning, and further improve the education of machine learning. Meanwhile, there may exist potential challenges that need to be addressed when implementing the proposed strategy, such as how to utilize the learnersourced annotations for training personalized classifiers and how to apply them to identify future biased learning materials. We hope that this paper offers some guidelines for potential studies on content moderation with learnersourcing.

3. Expected Implications and Broader Imapcts

Besides addressing the challenges in the current education strategy as described in Section 2.1, the designed content moderation strategy will also benefit the instructors, students, and the community in the long run. (1) **Improving the quality of course materials:** With the proposed course content moderation strategy, an individual student can provide their feedback about course materials with annotation, which further provides the opportunity for the instructors to improve the course materials to benefit the whole student population. We expect a more comprehensive, inclusive, and diverse set of course materials to cover various interests, backgrounds, and statuses of students. (2) **Actively involving students in learning:** The proposed strategy formulates the learning of classification algorithms as a classification task, which is easy to understand and motivates students to actively participate in the learning and further improve the learning outcome. (3) **Easy adaption to the education of other topics and disciplines:** We can easily adapt the proposed strategy to perform content moderation for other topics and disciplines, such as course materials about clustering and in the area of business. Moreover, besides the binary classification for course materials, we can also easily extend to the multi-class problems. (4) **Moderation of student-generated content:** We focused on the existing course materials in this paper, but the proposed strategy can be adapted to assess the student-generated content (e.g., submissions in homework assignments or course projects) before distributing to other students or utilizing in the future education. This will help the instructors to further maintain the quality of learnersourced materials.

We hope that by incorporating content moderation and other quality control strategies more meaningfully in learning activities, more diverse learning experiences can be created for students. In addition, content moderation learnersourcing has the potential to raise awareness and foster adequacy in creating ethical, responsible, and collegiate content for both students and instructors. By making content moderation meaningful, the proposed approach could encourage more engaged content moderation, curation, and quality control in existing learnersourcing solutions, and improve the quality of both existing and student-generated content. With a larger-scale adoption, the proposed method be used to develop the next generation of human-in-the-loop content monitoring tools for learnersourcing systems.

References

[1] P. Medel, V. Pournaghshband, Eliminating gender bias in computer science education materials, in: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, SIGCSE '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 411–416. URL: https://doi.org/10.1145/3017680.3017794. doi:10.1145/3017680.3017794.

[2] J. Kim, Improving learning with collective learner activity, 2015.

[3] L. L. Gaillet, A foreshadowing of modern theories and practices of collaborative learning: The work of scottish rhetorician george jardine., 1992.

[4] A. Singh, C. Brooks, S. Doroudi, Learnersourcing in theory and practice: Synthesizing the literature and charting the future, in: Proceedings of the Ninth ACM Conference on Learning @ Scale, L@S '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 234–245. URL: https://doi.org/10.1145/3491140.3528277. doi:10.1145/3491140.3528277.

[5] P. Denny, A. Luxton-Reilly, E. Tempero, J. Hendrickx, Codewrite: Supporting student-

driven practice of java, in: Proceedings of the 42nd ACM Technical Symposium on Computer Science Education, SIGCSE '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 471–476. URL: https://doi.org/10.1145/1953163.1953299. doi:10.1145/1953163.1953299.

[6] S. Downes, Connectivism and connective knowledge (2012).

[7] S. Myers West, Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms, New Media & Society 20 (2018) 4366–4383.

[8] D. Brown, The power and authority of materials in the classroom ecology, The Modern Language Journal 98 (2014) 658–661.

[9] A. Darvishi, H. Khosravi, S. Sadiq, Employing peer review to evaluate the quality of student generated content at scale: A trust propagation approach, in: Proceedings of the Eighth ACM Conference on Learning @ Scale, L@S '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 139–150. URL: https://doi.org/10.1145/3430895.3460129. doi:10.1145/3430895.3460129.

[10] O. Dele-Ajayi, J. Bradnum, T. Prickett, R. Strachan, F. Alufa, V. Ayodele, Tackling gender stereotypes in stem educational resources, in: 2020 IEEE Frontiers in Education Conference (FIE), IEEE, 2020, pp. 1–7.

[11] E. L. Glassman, A. Lin, C. J. Cai, R. C. Miller, Learnersourcing personalized hints, in: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work amp; Social Computing, CSCW '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1626–1636. URL: https://doi.org/10.1145/2818048.2820011. doi:10.1145/2818048.2820011.

[12] G. J. S. Dei, Decolonizing education for inclusivity: Implications for literacy education, Transcultural literacies: Re-visioning relationships in teaching and learning (2019) 5–30.

[13] E. B. King, L. M. Gulick, D. R. Avery, The divide between diversity training and diversity education: Integrating best practices, Journal of management education 34 (2010) 891–906.

[14] P. G. Devine, T. L. Ash, Diversity training goals, limitations, and promise: a review of the multidisciplinary literature, Annual review of psychology 73 (2022) 403.

[15] B. Xie, M. J. Davidson, B. Franke, E. McLeod, M. Li, A. J. Ko, Domain experts' interpretations of assessment bias in a scaled, online computer science curriculum, in: Proceedings of the Eighth ACM Conference on Learning@ Scale, 2021, pp. 77–89.

[16] B. K. Litts, K. A. Searle, B. M. Brayboy, Y. B. Kafai, Computing for all?: Examining critical biases in computational tools for learning, British Journal of Educational Technology 52 (2021) 842–857.

[17] A. Carvalho, Students' perceptions of fairness in peer assessment: evidence from a problem-based learning course, Teaching in Higher Education 18 (2013) 491–505.

[18] J. P. Sheldon, Gender stereotypes in educational software for young children, Sex Roles 51 (2004) 433–444.

[19] R. L. Blumberg, Gender bias in textbooks: A hidden obstacle on the road to gender equality in education, Unesco Paris, 2007.

[20] A. H. Kerkhoven, P. Russo, A. M. Land-Zandstra, A. Saxena, F. J. Rodenburg, Gender stereotypes in science education resources: A visual content analysis, PloS one 11 (2016) e0165037.

[21] G. Gezici, Y. Saygin, Measuring gender bias in educational videos: A case study on youtube,

arXiv preprint arXiv:2206.09987 (2022).

[22] S. Shaffer, L. Shevitz, She bakes and he builds: Gender bias in the curriculum, Double jeopardy: Addressing gender equity in special education (2001) 115–132.

[23] W. C. Sleeman IV, R. Kapoor, P. Ghosh, Multimodal classification: Current landscape, taxonomy and future directions, ACM Computing Surveys 55 (2022) 1–31.