Get Out of the BAG! Silos in AI Ethics Education: Unsupervised Topic Modeling Analysis of Global AI Curricula (Extended Abstract)

Rana Tallal Javed[1] , **Osama Nasir**[2] , **Melania Borit**[3] , **Loïs Vanhée**[4] , **Elias Zea**[5,7] , **Shivam Gupta**[6] , **Ricardo Vinuesa**[5,7] , **Junaid Qadir**[8]
[1]University of Oslo, Oslo, Norway
[2]Information Technology University of the Punjab, Pakistan
[3]UiT The Arctic University of Norway, Tromsø, Norway
[4]Umeå Universitet, Umeå, Sweden
[5]KTH Royal Institute of Technology, Stockholm, Sweden
[6]University of Bonnät Bonn, Bonn, Germany
[7]KTH Climate Action Centre, Stockholm, Sweden
[8]Qatar University, Doha, Qatar
ranaj@ifi.uio.no, msds19012@itu.edu.pk, melania.borit@uit.no, lois.vanhee@umu.se, zea@kth.se, shivam.gupta@uni-bonn.de, rvinuesa@mech.kth.se, jqadir@qu.edu.qa

Abstract

This study explores the topics and trends of teaching AI ethics in higher education, using Latent Dirichlet Allocation as the analysis tool. The analyses included 166 courses from 105 universities around the world. Building on the uncovered patterns, we distil a model of current pedagogical practice, the BAG model (Build, Assess, and Govern), that combines cognitive levels, course content, and disciplines. The study critically assesses the implications of this teaching paradigm and challenges practitioners to reflect on their practices and move beyond stereotypes and biases.

1 Introduction

With the explosive expansion of Artificial Intelligence (AI), teaching ethics to developers is being stressed increasingly by educational, governmental, and industrial organizations. This work analyzes the patterns of teaching AI ethics and critically assesses their implications.

Teaching AI and AI ethics and research on how this is done have been there for more than half a century [Chand, 1974; Gehman, 1984; Martin *et al.*, 1996; Applin, 2006; Ahmad, 2014], but the *systematic assessment* of the topics, developments, and trends in teaching AI ethics is a relatively recent endeavour. Previous research on a systematic analysis of teaching AI ethics suffered from many limitations: 1) *having a limited disciplinary scope* [Saltz *et al.*, 2019; Bielefeldt *et al.*, 2019; Khademi and Hui, 2020; Towell, 2003] 2) *having a limited geographical coverage* [Hughes *et al.*, 2020; Qadir and Suleman, 2018], 3) *being biased towards Western cultures* (e.g., [Moller and Crick, 2018], [Fiesler *et al.*, 2020], [Garrett *et al.*, 2020], [Raji *et al.*, 2021], [Homkes and Strikwerda, 2009]); or *including courses taught at only one single level* (e.g., introductory level [Becker and Fitzpatrick, 2019]).

Our analysis is based on unsupervised topic modelling, where we use the computational lens of the topic model to understand the AI ethics curricula. We use topic modelling to automatically uncover hidden or latent thematic structures from a textual corpus. The uncovered topics are derived from groups of co-occurring words (i.e. words that frequently come up together, within and between documents) that are associated with a single subject (or theme), which is referred to as a *topic* [DiMaggio *et al.*, 2013]. After identifying the hidden topics of AI ethics courses, we study the connection between these, the discipline of the department(s) giving the course, and core pedagogical concepts, i.e. Intended Learning Outcomes (ILOs) and their alignment with teaching modalities. In addition, we look at worldwide geographical trends, topic prevalence, and topic co-occurrence. Building upon this analysis, we distill a model of current pedagogical practices in the domain of teaching AI ethics. Our approach to analysing the use of pedagogical concepts in AI ethics courses is unique as, in contrast with previous research (e.g. [Fiesler *et al.*, 2020; Saltz *et al.*, 2019; Bielefeldt *et al.*, 2019]), we anchor our study in well-recognized canons from pedagogy science, as explained in Section 2.3.

2 Methods

2.1 Data Collection

Our initial data was taken from a curated list of tech ethics curricula by [Fiesler, 2018]. At the time of our analysis (January 2021), this contained 259 courses. We filtered out

The original paper is referenced as: Javed, R. T., Nasir, O., Borit, M., Vanhée, L., Zea, E., Gupta, S., Vinuesa, R. & Qadir, J. (2022). Get out of the BAG! Silos in AI Ethics Education: Unsupervised Topic Modeling Analysis of Global AI Curricula. Journal of Artificial Intelligence Research, 73, 933-965.

all the syllabi that were not in English or that did not relate to AI. Thus, we retained only 123 courses. We decided to make the data more global and added 43 additional AI ethics courses. Hence, the final analysis used a corpus of 166 syllabi from around the world. Each course was related to a discipline, derived from the teaching department listed on Fiesler's list and/or the course's webpage. We clustered all the departments into four categories: "computer science", "humanities", and "law", while the label "multidisciplinary" was given to courses associated with at least two departments.

2.2 Topic Modeling

In order to understand better the syllabi in our dataset, topic modelling was performed via Latent Dirichlet Allocation (LDA) [Blei *et al.*, 2003]. Our syllabi were in various formats: PDF, web pages (HTML), word documents, and text files. First, every syllabus format was turned into a text format. The text was then sanitized by removing stopwords and lemmatizing. In the end, `lda_mallet` was used to extract topics from the corpus. We used topic coherence scores to find the best number of topics.

Topics were named based on the hybrid content analysis method from [Baden *et al.*, 2020] by two experts in the ethics of AI and pedagogy, who assigned a label based on a close inspection of the 15 most probable words from each topic, the intertropical distance map, and content of the syllabi in the dataset.

2.3 Pedagogical Analysis

In this study, we used Bloom's taxonomy [Krathwohl, 2002] and Biggs' constructive alignment principle [Biggs and Tang, 2011] as the core of the pedagogical analysis. **Bloom's taxonomy** is a hierarchical model to classify educational learning objectives into levels of cognitive complexity: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate*, and *Create*. In curricula design, the taxonomy is for devising the course ILOs. The prevalence of a given Bloom cognitive level within a syllabus was assessed from the frequency in this syllabus of lexicographic indicators (specific verbs) related to each level. Similarly, for assessing the prevalence of teaching activities, we used the teaching modalities corresponding to each Bloom cognitive level as identified in [Wong *et al.*, 2019].

Biggs' constructive alignment principle states that the components in the teaching system, especially the teaching methods used and the assessment tasks, must be aligned with the learning activities assumed in the ILOs [Biggs and Tang, 2011]. In this study, we checked whether the words denoting teaching modalities are aligned with the words denoting a specific Bloom cognitive level (i.e. ILOs).

3 Key Results & Discussion

3.1 What is Taught and How

The 166 courses included in our analysis are taught at 105 universities around the world. We categorized the courses based on their level (i.e., undergraduate or graduate) and the department associated with the course. Our results indicate that the instances of the Law department teaching AI ethics are fewer compared to the instances of AI ethics courses taught by the Humanities and Computer Science departments. Moreover, the geographical analysis shows that **the continent has an impact on the discipline of the department delivering the course**. This difference is important because various disciplines might train for various types of professional groups (e.g., engineers, managers, scientists). A major difference among societies with regard to what department delivers AI ethics education can lead to different outcomes in terms of the operational capacity societies will have in the future for dealing with AI ethics matters and, therefore, how AI ethics is to be implemented differently by these societies. For example, a society where AI ethics education is delivered mainly by technical departments might foster the emergence of a techno-centric approach to solving AI ethics issues.

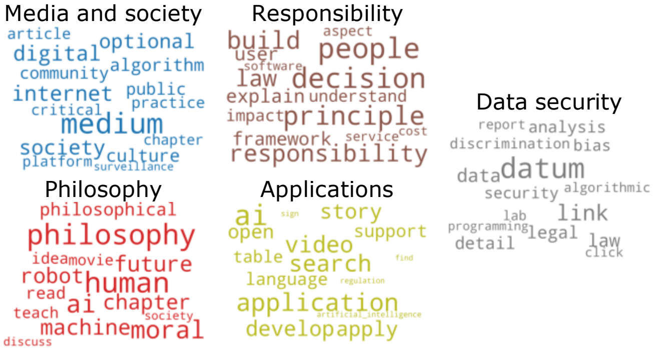

Figure 1: Main (non-administrative) topics identified by the LDA analysis applied on the corpus of the texts of AI ethics syllabi and most prevalent terms per topic.

The LDA analysis uncovered 10 topics: five topics related to course administration and five to course content: **T1** *Media & society*, **T4** *Philosophy*, **T6** *Responsibility*, **T8** *Data security*, and **T9** *Applications* (Figure 1). In order to focus the analysis on the actual course content and avoid being sidetracked by the administrative content, the administrative topics were hidden in the following analysis, and the remaining topics were normalized.

The heatmap displayed in Figure 2 shows the cooccurrence among topics since some topics tend to be used together in syllabi more frequently than others. This heatmap shows, given a dominant topic (set by the label of the row), the distribution of the prevalence of remaining topics. Analyzing the potential relations regarding the dominant topic and the prevalence of other topics, it is interesting to note that many of these relationships are asymmetrical.

Figure 3 answers the question: "*What is the prevalence of the various topics depending on the discipline of the department associated with the syllabus?*" This figure exhibits a form of discipline-based specialization.

Figure 4 answers the question: "*What is the prevalence of the various Bloom cognitive levels for each given topic?*" This analysis matches *a-priori* expectations one may have when relating a topic to a Bloom cognitive level.

Figure 5 answers the question: "*What differentiates the syllabus from different departments in its formulation for each Bloom cognitive levels?"* These results seem to confirm the stereotype that computer scientists are solution-oriented

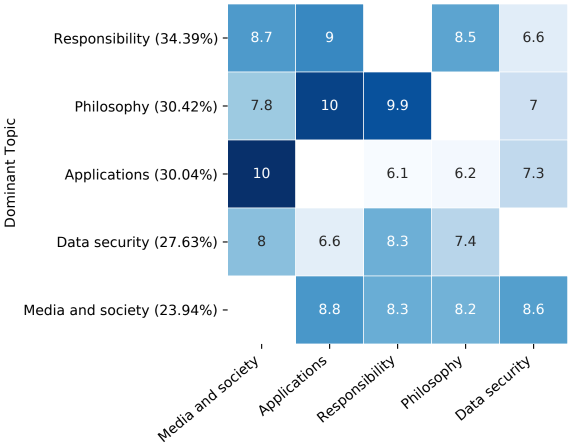

Figure 2: Heat map displaying the dominant topic (left) and the remaining average topic proportions (top) for all 166 syllabi,indicating the extent to which documents about one main topic relate to the other uncovered topics. For example, documents which main topic is *Responsibility* also co-occur with *Applications* (9%) or *Media & society* (8.7%).

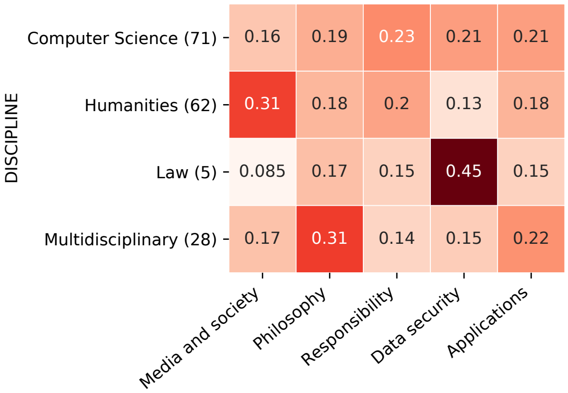

Figure 3: Distribution of the various topics depending on the discipline. Each row is normalized to 1. Number of offered courses is displayed in brackets.

while those coming from humanities and law focus most on remembering and basic comprehension of facts and concepts.

Figure 6 answers the question: "*What is the prevalence of the various topics in the syllabi of each continent?*" To our knowledge, this analysis of what topics in AI ethics are taught across the globe is the first one of this kind.

Topic	Number of occurrences
Responsibility	77 (38.5%)
Media & society	37 (18.5%)
Data security	37 (18.5%)
Philosophy	25 (12.5%)
Applications	24 (12%)

Table 1: Number of occurrences in the ACM ethics guidelines of the keywords of the various topics uncovered by the LDA.

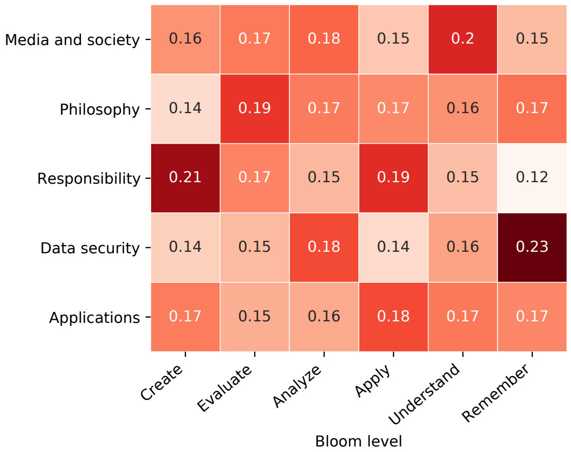

Figure 4: Each row represents the distribution of topics with respect to Bloom's cognitive levels. Rows are normalized to 1.

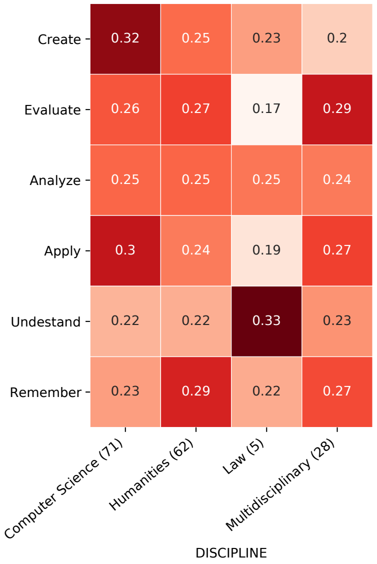

Figure 5: Distribution of Bloom cognitive levels depending on the department delivering the course (rows normalized to 1).

As a means for assessing the pedagogical content of a syllabus, we also analysed the teaching modalities (i.e., what teaching activities are to be performed to achieve the ILOs), as mentioned in the syllabus text. The results indicate recurrent misalignments between ILOs and teaching modalities across continents and departments.

For assessing which of the topics uncovered by the LDA are best aligned with concrete ethical concerns, we compared the keywords linked to the identified LDA topics with the contents of the ACM ethics guidelines. As an evaluation criterion, for every topic, we summed the number of occurrences of each of its keywords within the ACM ethics guidelines. This analysis highlights imbalances between the main approaches for teaching AI ethics and the representativeness of the taught concepts within the ACM ethics guidelines (see

	Build	Assess	Govern
Topic	*Responsibility*, *Applications*	*Philosophy*, *Applications*	*Media & society*, *Data security*
Discipline	Computer Science	Multidisciplinary	Humanities & Law
Bloom cognitive level	*Create*, *Apply*	*Evaluate*, *Analyze*	*Understand*, *Remember*

Table 2: A meta-synthesis of the approaches for teaching tech ethics: the BAG model (Build, Assess, Govern). The *Applications* topic bridges both the Build and the Assess teaching approaches.

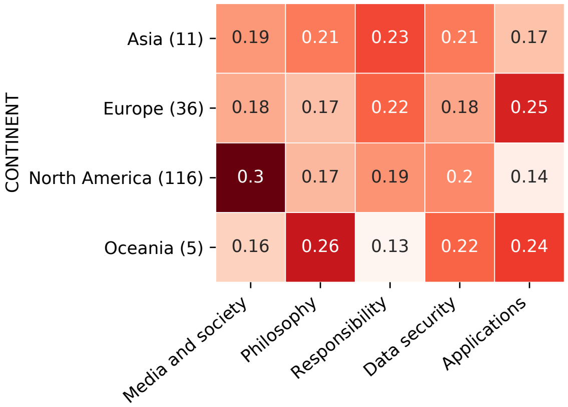

Figure 6: Distribution of the various topics depending on the continent. Each row is normalized to 1.

Table 1).

3.2 In the BAG and Out

A model of how AI ethics is taught across the world can be derived from a meta-synthesis of the trends analysis results. This model, i.e. the **BAG model** (for **Build, Assess, Govern**), consists of three general approaches, each of these approaches relating specific Bloom cognitive levels, course content topics, and disciplines associated with a syllabus, as summarized in Table **??**.

The BAG model shows that ethics education is still taught in isolated ways: specific disciplines, topics, and levels of ability. This is despite the general agreement that ethical and responsible AI design requires an interdisciplinary approach. Tomorrow's engineers are now trained with a focus on design but with limited insight into the legality or social desirability of their systems. Tomorrow's lawyers are now trained to see the risks and apply rules, with only minimal training on pragmatically weighing these risks over social benefits. Tomorrow's social scientists are now trained to see how AI can influence the dynamics of society, with only limited oversight on actual technical and engineering intricacies that can drive design decisions. This development in silos highlights an important issue that calls for action as a community if we want future generations to work together rather than against each other and to prevent society from becoming locked into undesirable path dependencies. Practitioners interested in renewing their practices could use the BAG model to identify where they position themselves relative to the three teaching approaches and try to cross topic bridges (e.g., *Data security* + *Responsibility*), Bloom cognitive level bridges (*Remember* + *Apply*), or discipline bridges (Humanities + Computer Science). As a matter of removing these silos and getting out of the BAG model, it seems relevant to strive for a form of holistic training, removing disciplinary, topic-oriented, and Bloom cognitive levels barriers

4 Conclusions

This study uses advanced statistical tools, such as a Latent Dirichlet Allocation (LDA) analysis and a meta-synthesis, to explore what is taught in AI courses around the world and how. It follows well-known canons from pedagogy science, i.e. Bloom's taxonomy and Biggs' constructive alignment principle. Unlike previous studies, our analysis covers a wide range of disciplines, regions (global level), and levels of education (undergraduate and graduate), and it uses automated statistical methods. The analysis indicates that there are numerous significant misalignments between the ILOs formulated for the courses included in our corpus and the teaching modalities put in place for training these ILOs. Furthermore, the analysis highlights the presence of three silos related to the pedagogy strategies for teaching AI ethics along a model which we call BAG (Build, Assess, Govern). The findings of this study highlight that current curricula perpetuate disciplinary mindsets and communities, which is suboptimal for the seamless design of systems that best serve society. This study suggests a solution: a generic-to-specific hybrid teaching and learning approach that connects different communities with a common understanding of the value of AI ethics.

Acknowledgments

LV acknowledges the support of the Knut and Alice Wallenberg Foundation (project number 570080103) and of the project AutogrAIde "A Student-Driven Interdisciplinary Hackathon on Whether and How to Automate Grading & Assessment" (project number 570002260, The Rådet för Artificiell Intelligens, Umeå University, Sweden). MB acknowledges the support of the project AFO-JIGG "Service design thinking to improve welfare and product quality in the Norwegian small-scale fishing fleet" (project number 302635, Research Council of Norway). JQ acknowledges the support of the project from Qatar University High-Impact Grant QUHI-CENG-23/24-127. SG acknowledges the funding provided by the German Federal Ministry for Education and Research (BMBF) for the project "digitainable".

References

[Ahmad, 2014] Farooq Ahmad. Computer science & engineering curricula and ethical development. In *2014 International Conference on Teaching and Learning in Computing and Engineering*, pages 220–225. IEEE, 2014.

[Applin, 2006] Anne G Applin. A learner-centered approach to teaching ethics in computing. *ACM SIGCSE Bulletin*, 38(1):530–534, 2006.

[Baden *et al.*, 2020] Christian Baden, Neta Kligler-Vilenchik, and Moran Yarchi. Hybrid content analysis: Toward a strategy for the theory-driven, computer-assisted classification of large text corpora. *Communication Methods and Measures*, 14(3):165–183, 2020.

[Becker and Fitzpatrick, 2019] Brett A Becker and Thomas Fitzpatrick. What do cs1 syllabi reveal about our expectations of introductory programming students? In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, pages 1011–1017, 2019.

[Bielefeldt *et al.*, 2019] Angela R Bielefeldt, Madeline Polmear, Daniel Knight, Nathan Canney, and Christopher Swan. Disciplinary variations in ethics and societal impact topics taught in courses for engineering students. *Journal of Professional Issues in Engineering Education and Practice*, 145(4):04019007, 2019.

[Biggs and Tang, 2011] John B Biggs and C. Tang. *Teaching for quality learning at university*. McGraw-hill education (UK), 2011.

[Blei *et al.*, 2003] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[Chand, 1974] D R Chand. Computer science education in business schools. In *ACM Special Interest Group on Computer Science Education Bulletin*, volume 6, pages 91–97, 1974.

[DiMaggio *et al.*, 2013] Paul DiMaggio, Manish Nag, and David Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6):570–606, 2013.

[Fiesler *et al.*, 2020] Casey Fiesler, Natalie Garrett, and Nathan 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*, pages 289–295, 2020.

[Fiesler, 2018] Casey Fiesler. Tech Ethics Curricula: A Collection of Syllabi. https://tinyurl.com/msuh396p, 2018. Accessed: 2021-10-10.

[Garrett *et al.*, 2020] Natalie Garrett, Nathan Beard, and Casey Fiesler. More than “if time allows”: The role of ethics in AI education. In *Proceedings of 2020 AAAI-ACM Conference on Artificial Intelligence, Ethics, and Society (AIES’20)*, pages 272–278, 2020.

[Gehman, 1984] William G Gehman. Responsible Software Engineering. In *ACM 1984 Comp. Sci. Conf*, page 177, 1984.

[Homkes and Strikwerda, 2009] Rick Homkes and Robert A Strikwerda. Meeting the ABET program outcome for issues and responsibilities: an evaluation of CS, IS, and IT programs. In *Proceedings of the 10th ACM conference on SIG-information technology education*, pages 133–137, 2009.

[Hughes *et al.*, 2020] Janet Hughes, Ethan Plaut, Feng Wang, Elizabeth von Briesen, Cheryl Brown, Gerry Cross, Viraj Kumar, and Paul Myers. Global and local agendas of computing ethics education. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, pages 239–245, 2020.

[Khademi and Hui, 2020] Keyvan Khademi and Bowen Hui. Towards understanding the HCI education landscape. In *Koli Calling’20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*, pages 1–2, 2020.

[Krathwohl, 2002] David R. Krathwohl. A Revision of Bloom’s Taxonomy: An Overview. *Theory Into Practice*, 41(4):212–218, 2002.

[Martin *et al.*, 1996] C. Dianne Martin, Chuck Huff, Donald Gotterbarn, and Keith Miller. A framework for implementing and teaching the social and ethical impact of computing. *Education and Information Technologies*, 1(2):101–122, 1996.

[Moller and Crick, 2018] F. Moller and T. Crick. A university-based model for supporting computer science curriculum reform. *J. Comput. Educ.*, 5:415–434, 2018.

[Qadir and Suleman, 2018] J. Qadir and M. Suleman. Teaching ethics, (islamic) values and technology: Musings on course design and experience. In *2018 7th International Conference on Computer and Communication Engineering (ICCCE)*, pages 486–491, 2018.

[Raji *et al.*, 2021] Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. You can’t sit with us: Exclusionary pedagogy in AI ethics education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 515–525, 2021.

[Saltz *et al.*, 2019] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. Integrating ethics within machine-learning courses. *ACM Transactions on Computing Education*, 19:1–26, 2019.

[Towell, 2003] Elizabeth Towell. Teaching ethics in the software engineering curriculum. In *Proceedings 16th Conference on Software Engineering Education and Training, 2003.(CSEE&T 2003).*, pages 150–157. IEEE, 2003.

[Wong *et al.*, 2019] Mun Kit Wong, Jiaxuan Wu, Zhi Yang Ong, Jia Ling Goh, Clarissa Wei Shuen Cheong, Kuang Teck Tay, Laura Hui Shuen Tan, and Lalit Kumar Radha Krishna. Teaching ethics in medical schools: A systematic review from 2000 to 2018. *Journal of Medical Education*, 18:226–250, 2019.