Journal of Artificial Intelligence Research 73 (2022) 933-965 Submitted 12/2021; published 03/2022

Get out of the BAG! Silos in AI Ethics Education: Unsupervised Topic Modeling Analysis of Global AI Curricula

Rana Tallal Javed tallal.javed@itu.edu.pk
Osama Nasir msds19012@itu.edu.pk
Information Technology University of the Punjab
Electrical engineering department
Arfa Software Technology Park
Ferozepur Road, 142036 Punjab, Pakistan

Melania Borit melania.borit@uit.no
UiT The Arctic University of Norway
Norwegian College of Fisheries Science
Muninbakken 21, 9019 Tromsø, Norway

Loïs Vanhée *(Corresponding author)* lois.vanhee@umu.se
Umeå Universitet
Department of Computing Science
Linnaeus väg 49, 907 36 Umeå, Sweden

Elias Zea zea@kth.se
KTH Royal Institute of Technology
Department of Engineering Mechanics
Teknikringen 8, 100 44 Stockholm, Sweden

Shivam Gupta shivam.gupta@uni-bonn.de
Bonn Alliance for Sustainability Research
University of Bonn
Regina-Pacis-Weg 3, 53113 Bonn, Germany

Ricardo Vinuesa rvinuesa@mech.kth.se
KTH Royal Institute of Technology
FLOW, Engineering Mechanics
Osquars backe 18, 100 44 Stockholm, Sweden

Junaid Qadir jqadir@qu.edu.qa
Computer Science and Engineering Department
Qatar University
2713 Doha, Qatar

Abstract

The domain of Artificial Intelligence (AI) ethics is not new, with discussions going back at least 40 years. Teaching the principles and requirements of ethical AI to students is considered an essential part of this domain, with an increasing number of technical AI courses taught at several higher-education institutions around the globe including content related to ethics. By using Latent Dirichlet Allocation (LDA), a generative probabilistic topic model, this study uncovers topics in teaching ethics in AI courses and their trends related to where the courses are taught, by whom, and at what level of cognitive complexity and specificity according to Bloom's taxonomy. In this exploratory study based on unsupervised machine learning, we analyzed a total of 166 courses: 116 from North American universities, 11 from Asia, 36 from Europe, and 10 from other regions. Based on this analysis, we were able to synthesize a model of teaching approaches, which we call BAG (Build, Assess, and Govern), that combines specific cognitive levels, course content topics, and disciplines affiliated with the department(s) in charge of the course. We critically assess the implications of this teaching paradigm and provide suggestions about how to move away from these practices. We challenge teaching practitioners and program coordinators to reflect on their usual procedures so that they may expand their methodology beyond the confines of stereotypical thought and traditional biases regarding what disciplines should teach and how.

1. Introduction

Artificial Intelligence (AI) is advancing at an explosive rate, playing a significant role in how we communicate, learn, and interact in society. AI professionals are creating intelligent solutions for almost every domain, ranging from healthcare and food processing to entertainment and even warfare. More voices than ever recognize the opportunities offered by AI, while also insisting on facilitating a responsible, ethics-driven, inclusive, and context-aware implementation (Floridi et al., 2018)[1]. The AI community as a whole seems to be more aware than ever of the ethical crisis unleashed by AI (Tzachor et al., 2020).

An essential solution stressed for decades by educational, governmental, and industrial organizations alike for addressing problematic issues that are raised or might be raised by technological development has been to incorporate ethics into teaching AI to tech professionals and to include in the curricula of future AI practitioners specific training that ranges from raising ethical awareness to developing concrete skills for the implementation of ethical guidelines (Nielsen, 1972; Goldsmith & Burton, 2017; Dignum, 2021). This approach has been also acknowledged by, for example, the Special Interest Group in Computer Science Education (SIGCSE)[2], constituted under the Association for Computing Machinery (ACM), which has been organising a technical symposium annually since 1970, during which participants address problems common among educators and work to develop, implement and/or evaluate computing programs, curricula, and syllabi.[3]

1. Some academics have also started the debate about how ethics-driven AI can also contribute to sustainable development (Vinuesa et al., 2020; Henriksson & Grunewald, 2020; Gupta et al., 2020). Some of the areas where ethics-driven AI can have a positive impact on sustainability are related with interpretable and transparent deep-learning models. As discussed in the literature (Vinuesa & Sirmacek, 2021), certain applications aimed at tackling the United Nations Sustainable Development Goals (UNSDGs) related to eradication of poverty (Jean et al., 2016) may significantly benefit from interpretable models, due to the fact that more profound actions can be taken to reduce poverty when the predictive model is transparent. Other important areas are related to equality and diversity of the AI working force, which will have positive effects on the UNSDGs related to gender equality and reduced inequalities.
2. https://sigcse.org/events/symposia/index.html
3. The domain of teaching AI ethics is rapidly advancing, with more than 180 academic articles published on this topic since 2020 (according to a search in Scopus, December 2021). Many of these studies discuss more general aspects of AI ethics education (e.g., Hwang et al., 2020; Lin et al., 2021; Yang et al., 2021), while many others explore effective ways to teach the desired topics and skills (e.g., the use of science fiction Burton, Goldsmith, & Mattei, 2018 or design fiction / speculative design Soden et al., 2019 to help anticipate and reflect on the potential downsides of technology design and to help recognizing potentially harmful scenarios). Practitioners have also built dedicated arenas for teaching tech ethics, for example, Embedded EthiCS at Harvard (Grosz et al., 2019); the Mozilla Foundation (`https://foundation.`

Higher-education institutions have been trying to find the best ways in which to effectively teach the necessary knowledge, skills, and competences that can effectively and holistically train students (i.e., future technical AI developers) to blend an ability for AI development and use with an ethical mindset. The groundswell of interest around the theme of teaching AI ethics can be seen in the long list of recent tech ethics courses collected in a crowd-sourced list available at `http://tinyurl.com/ethics-classes`. The purpose of the present study is to analyze *patterns of teaching AI ethics and to critically assess their implications*.

While academics have been studying the topics of teaching AI and AI ethics for more than half a century (e.g., Chand, 1974; Gehman, 1984; Martin et al., 1996; Applin, 2006; Ahmad, 2014), the *systematic assessment* of the topics, developments, and trends in teaching AI ethics is a relatively recent endeavor. However, most of the previous research that focused on a systematic analysis of teaching AI ethics suffered from one or more of the following limitations: 1) *having a limited disciplinary scope* (e.g., integration of ethics only in courses in machine-learning, Saltz et al., 2019; engineering, Bielefeldt et al., 2019; Nasir et al., 2021; human-computer interaction, Khademi & Hui, 2020; software engineering, Towell, 2003; or distributed systems, Abad, Ortiz-Holguin, & Boza, 2021); 2) *having a limited geographical coverage* and, as explained in Hughes et al. (2020), Mohamed et al. (2020), *being biased towards Western cultures* (e.g., Moller & Crick, 2018; Fiesler et al., 2020; Garrett et al., 2020; Raji et al., 2021; Homkes & Strikwerda, 2009); or 3) *including courses taught at only one single level* (e.g., introductory level, Becker & Fitzpatrick, 2019).

Most importantly, all these previous attempts to map the teaching AI ethics field are human-driven approaches, with topics of interest manually identified based on grouping the instructor-described topics into higher-level categories (e.g., Fiesler et al., 2020) or on open coding (e.g., Garrett et al., 2020; Raji et al., 2021). However, such approaches are sensitive to the subjectivity and noise inherent in human decisions and the limited ability of human analysts to work effectively at very large scales. To complement the previous approaches and expand the understanding of the situation, this study pursues a novel strategy for the field of AI ethics education and leverages an automated data-driven approach based on topic modelling (Debortoli et al., 2016) to uncover hidden topics within AI ethics courses. More specifically, we uncover these topics and their trends using a unique dataset consisting of 166 syllabi from AI ethics courses around the world. We do that by using hybrid content analysis, which includes the analysis of content combining the algorithmic extraction of coherent and recurrent patterns with human interpretation of identified patterns (Baden, Kligler-Vilenchik, & Yarchi, 2020) (details in Section 2).

Topic modelling relies on machine-learning methods for automatically uncovering hidden or latent thematic structures from a textual corpus. The uncovered topics are derived from groups of co-occurring words that are associated with a single subject (or theme), which is referred to as a *topic* (DiMaggio, Nag, & Blei, 2013). Co-occurring words are words that tend to appear together within the same linguistic context more frequently than one would expect by chance alone. In a nutshell, topic models are able to exploit the co-occurrence structure of texts and produce the topics as lists of words that frequently come up together,

`mozilla.org/en/what-we-fund/awards/teaching-responsible-computing-playbook/topics/`; or the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) program (`https://wasp-hs.org/projects/`).

within and between documents. Technically, such lists of words are probability distributions over words, for example, garden plants = {0.2: roses, 0.4: lilies, 0.0: AI, 0.3: spring, 0.2: autumn}. For example, in the context of an AI ethics educational program, course designers might use the following words in the course titles: "programming", "economics", "ethics", "machine learning", "AI". However, the syllabus might use more frequently the words "law", "principles", "responsibility", "decision", "data", "bias", "security", "discrimination". If we wanted to use topic modelling to uncover the latent topics of this hypothetical program, based on how often these most used words would appear together (i.e., co-occur), the automated topic model would group the first four words into one topic and the last four words into a different topic. Because the topic-modelling algorithm does not assign a *subject* (i.e., a label) to the uncovered topics, these two topics would then be manually labelled by a domain expert, as the manual interpretation of the *subject* of a topic is the most straightforward approach of most studies regarding valid interpretation of the resulting topics (Maier et al., 2018). In our example, the two topics would most likely be labelled as *Responsible design* and *Data security* respectively. Note that the subject of these two topics determined based on the topic modelling analysis is not similar to the one that might be inferred from the titles given to the courses included in this hypothetical program. Thus, these topics are latent, and they are hidden in the pattern of co-occurring words.

Generally speaking, topic modelling approaches have been helpful in investigating the key ideas within a set of documents, such as articles published in academic journals (e.g., marine sciences journals Syed, Borit, & Spruit, 2018), political texts (Grimmer & Stewart, 2013), or data-driven journalism (Rusch et al., 2013). In the domain of AI/Computer Science (CS), topic modelling has been used to identify salient topics discussed in national AI policies (van Berkel et al., 2020), topics prevalent in Internet news related to the fourth industrial revolution (Jang, Park, & Kim, 2018), or topics of research on AI in marketing (Mustak, Salminen, Plé, & Wirtz, 2021). The method has also been used recently in the CS education domain for detecting latent topics and trends in educational technologies (Chen et al., 2020), improving teaching material (Marçal et al., 2020), or analyzing the curriculum of computer science departments (Matsuda, Sekiya, & Yamaguchi, 2018; Sekiya, Matsuda, & Yamaguchi, 2010).

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 derived from Bloom's taxonomy (Krathwohl, 2002), which, according to Fuller, Johnson, Ahoniemi, Cukierman, Hernán-Losada, Jackova, Lahtinen, Lewis, Thompson, Riedesel, and Others (2007), dominates the field of CS course and assessment design. In addition, we look at worldwide geographical trends, topic prevalence, and topic co-occurrence. After doing that, we distil 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, that is, Bloom's taxonomy (Krathwohl, 2002) and Biggs' constructive alignment principle (Biggs & Tang, 2011), as explained in Section 2.3.

Systematized approaches as topic modelling can complement pure human annotation methods, as follows. 1) It can reduce noise in processing large amounts of text (Kahneman, Sibony, & Sunstein, 2021). 2) The analysis output generates topics that are identified based on statistical regularities of the words, thus the topics match the words, something which is difficult to achieve when the analysis is done by hand, particularly with large amounts of text. 3) The analysis output provides the relative prevalence of each topic within each item included in the corpus, whereas such a fine-grained relation is difficult, if not impossible, to produce with manual coding. 4) The analysis automates a large part of the process, allowing for efficient re-iteration of the data-processing activities. In the long run, particularly in a context where the number of AI ethics courses is increasing at a very fast pace, this automated process will allow assessing how the contents of the courses are evolving over time with greater speed and quantitative rigour than would otherwise be possible through traditional reviews (Grimmer & Stewart, 2013). However, as any other method, LDA is subject to bias, notably arising from, as explained in 3.4, the bias originating from the raw data (Gitelman, 2013), the generalizations made by the algorithm (Mitchell, 1980), and human topic labelling.

2. Methodology

2.1 Data Collection

Syllabi Data: In late 2017, a controversy started with a New York Times op-ed declaring the academics community to be "asleep at the wheel" concerning tech ethics (O'Neil, 2017). This controversy stoked a trend that motivated researchers to investigate this claim and led to the creation of a community-based public compilation of Tech Ethics curricula in the form of a crowd-sourced Google sheet (Fiesler, 2018). This sheet provides a list of tech ethics courses, referencing meta-information including course title, course instructor, course level (undergraduate or graduate), the teaching department, university, and the possibility to include links towards course descriptions and curricula. At the time of our analysis (January 2021), it contained 259 courses. Since our aim was to analyze the syllabi rather than the metadata, we filtered out all the courses that did not have their syllabus publicly available. Then, we filtered out all the syllabi that were not in English or that did not relate to AI, by systematically removing syllabi that did not contain words such as AI, machine learning, deep learning, or data science. As a result, a total of 123 courses were retained from this Google Sheet and the rest were excluded.

To make the dataset more global, we decided to expand this list, which initially contained mostly courses located in the United States of America (US) and Europe. Therefore, special focus was given to identifying courses from the under-represented regions. We used a list of search strings along with various country/continent names to find similar courses (see Appendix A). A course was added to the dataset if it was related to AI ethics (as described above) and had an accessible syllabus in English. This activity led us to expand the dataset with *43 additional AI ethics courses*. As a result, a total of 166 courses were included in the final dataset and were subsequently analyzed. Metadata of the final dataset are shared as online Supplementary material. The syllabi of all these courses were downloaded and included in the analysis.

Each course is related to a *discipline*, derived from the teaching department listed on the Fiesler's list and/or the webpage of the course. We clustered all the departments in four categories. All the departments that were related to engineering, computer science, informatics etc. were categorized as "*computer science*". Humanities related departments, such as communication, media studies etc., were clustered under "*humanities*", and the law and policy related departments were grouped under "*law*". The label "*multidisciplinary*" was given to courses associated with at least two departments, that is, it was being taught together by two or more departments.

Ethical Considerations: No personal information, like emails or contact numbers, was retained in the analysis. All the used curricula were publicly available; they were obtained from official websites and were analyzed "as is". In our analysis, we ensured that the identity of individual subjects was protected and that there was no risk of harm to the individuals and organizations that made these data publicly available due to the research presented in this paper, in accordance with the guidelines by Markham and Buchanan (2012).

2.2 Topic Modelling

2.2.1 Latent Dirichlet Allocation

In order to understand better the syllabi in our dataset, topic modelling was performed via Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003). LDA is a generative probabilistic model widely used in the Natural Language Processing community, with applications ranging from document query, to text classification, and topic modelling. The LDA representation has three levels: corpus, document (syllabi text, in our case), and topic. The main idea is to represent documents with random mixtures over a set of latent topics, which, in turn, are represented with mixtures over a set of topic probabilities. Further technical details can be found in 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. Then, standard stopwords, detailed in Annex B, were removed. Next, words were lemmatized using `spaCy`[4] and nouns, verbs, adjectives, and adverbs were kept. Then, topics and keywords per topics were extracted from the corpus using `Gensim`'s[5] LDA implementation. The model used was `lda_mallet`.

2.2.2 Parameter Optimization

Topic modeling is sensitive to one hyper-parameter, that is, the number of considered topics, which is critical to optimize. There are both intrinsic and extrinsic measures to optimize the number of topics identified by the LDA. The intrinsic score checks the coherence between generated topics based on the word co-occurrence within the generated corpus. This measure checks the coherence of topics with respect to the documents used for training the model. The extrinsic score uses some external pre-defined corpus to compute the topic coherence. These measures are usually employed where generated topics are evaluated, and their coherence is tested with respect to language in general. "u_mass"[6] was used for computing the score as an intrinsic measure to compute topic coherence. "u_mass" uses a pairwise score function for a pair of words (w_i, w_j), which is the smoothed conditional log probability $\log p(w_j|w_i)$ over the corpus.

4. `https://spacy.io/api/lemmatizer`
5. `https://radimrehurek.com/gensim/`
6. `https://radimrehurek.com/gensim/models/coherencemodel.html`

2.2.3 Labelling Topics

The topical structure that characterises the dataset is latent and the probability distributions of words (i.e., topics) are not given a name by the LDA model. When sorted, usually, the top 10-15 most probable words within each topic are used to describe that topic. The manual interpretation of the *subject* of a topic is the most straightforward approach of most studies regarding valid interpretation of the resulting topics (Maier et al., 2018). Following the hybrid content analysis method described in Baden et al. (2020), in our study, giving a name to each topic was performed by human analysts (i.e., two experts in ethics of AI and pedagogy). First, the analysts closely inspected the 15 most probable words from each topic. Second, the analysts inspected the intertopic distance map and the titles and content of the syllabi in the dataset that were included by the topic model in that respective topic. Third, based on the previous two steps, the analysts gave a label to each of the topics.

2.3 Pedagogical Analysis

In this study, we use Bloom's taxonomy (Krathwohl, 2002) and Biggs' constructive alignment principle (Biggs & Tang, 2011) at the core of the pedagogical analysis.

Bloom's taxonomy is a hierarchical model to classify educational learning objectives into levels of cognitive complexity: from basic to increasingly complex knowledge. In its original form (Bloom et al., 1965), the taxonomy consisted of six cognitive levels: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate*, and *Synthesize*. The revised form (Krathwohl, 2002) essentially differs in the last level: *Create*. In course curriculum design, the taxonomy is broadly used for devising the Intended Learning Outcomes (ILOs), alternatively called Course Learning Outcomes (CLOs) in some outcome-based education and accreditation regimes (Qadir et al., 2020), and for choosing the right teaching modalities or activity (e.g., lecture, group work, project work, internship).

Bloom's cognitive levels can be summarized as follows. *Remember* refers to recalling facts, terms, basic concepts, and answers. *Understand* refers to basic comprehension of facts and ideas. *Apply* refers to solving problems by using acquired knowledge, facts, techniques, and rules, in a different way. *Analyze* refers to examining and breaking information into parts by identifying motives or causes, making inferences, and finding evidence to support generalisations. *Evaluate* refers to presenting and defending opinions by making judgements about information, validity of ideas or quality of something based on a set of criteria. *Create* refers to compiling information together in a different way, by combining elements in a new pattern, or proposing an alternative solution.

As a means for assessing the pedagogical content of a syllabus, we used two criteria: *the Bloom cognitive level* (i.e., what ILOs are to be achieved), as indicated by the syllabus through the verbs used, and *the teaching modalities* (i.e., what teaching activities are to be performed to achieve the ILOs), as mentioned in the syllabus text. For assessing the prevalence of a given Bloom cognitive levels within a syllabus, every Bloom cognitive level was related to a set of level-specific verbs taken from Bloom's taxonomy[7] listed in Annex C. 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). The prevalence of a given Bloom cognitive level / teaching modality for a given syllabus was computed

7. `https://tips.uark.edu/blooms-taxonomy-verb-chart/`

as the frequency of occurrence of the words related to the considered Bloom cognitive level / teaching modality in the syllabus. For example, if five occurrences of *Evaluate* Bloom cognitive level verbs are identified in a 250-word long curriculum c, then the proportion of the *Evaluate* Bloom cognitive level in c is defined as $5/250 = 0.02$. The higher this number, the more prevalent is the Bloom cognitive level/ teaching modality.

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 intended outcomes (Biggs & 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); for example, if words like lecture, seminar, test, as teaching modalities for *Remember* Bloom cognitive level co-occur with words like identify, indicate, list, cite, define etc. for *Remember* Bloom cognitive level.

3. Results and Discussion

3.1 Metadata Analysis

During the analysis of the meta-information of the 166 courses, we categorized the courses based on their level (i.e., undergraduate or graduate) and the department associated with the course (Table 1). Apart from this, we also studied the geographical distribution of these courses (Figure 1). A total of 105 universities are part of our dataset, with many top universities offering multiple courses. Some syllabi focus on understanding the ethical aspects related to the AI industry, while others focus on developing a sense of ethics and responsibility in CS students. 94 syllabi (∼56.6%) describe courses offered to undergraduate students, 49 syllabi (∼29.5%) describe courses offered to graduate students, and 23 syllabi (∼13.9%) describe course offered to both graduate and undergraduate students.

The departmental and geographical breakdown of these courses is shown in Table 1. It can be observed 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. One potential explanation for this imbalance can be that Law departments do not extensively use AI related words in the title of their courses, thus the probability of such courses to be identified by search engines decreases. In addition, because of the conceptual differentiation made between ethics and law, teachers from Law departments might not consider registering on the tech ethics list from Fiesler (2018).

An analysis of the distribution of the departments delivering the courses depending on the continent shows that the continent has an impact on the type of department delivering the course. In Asia, most of the AI ethics courses are delivered by the Computer Science departments. In Europe, AI ethics courses are offered by the Computer Science departments, followed by a multidisicplinary offer, and then, to a less extent, by the Humanities departments. In North America, Humanities departments are the main deliverer of AI ethics courses, followed by Computer Science, and then, followed at a large distance, by a multidisciplinary offer. In Oceania and South America, most of the AI ethics courses are offered by Computer Science departments and to a less extent by the Humanities departments. Further, it can be observed that courses delivered by Law departments are only present in North America and multidisciplinary offers are only observed in Europe and North America. Comparing the most represented continents (i.e., Europe and North America), it can be ob-

Table 1: Department-wise disciplinary distribution of AI ethics courses included in our corpus. Multidisciplinary courses are considered to be those courses offered by at least two different departments.

Discipline	Courses	Couses/Continent		Intra-continental proportion
Computer Science	71 (42.8%)	Asia:	9	81.8%
		Europe:	15	45.5%
		North America:	43	37.1%
		Oceania & South America:	4	66.7%
Humanities	62 (37.3%)	Asia:	2	18.2%
		Europe:	6	18.2%
		North America:	52	44.8%
		Oceania & South America:	2	33.3%
Multidisciplinary	28 (16.9%)	Asia:	0	0%
		Europe:	12	36.4%
		North America:	16	13.8%
		Oceania & South America:	0	0%
Law	5 (3.0%)	Asia:	0	0%
		Europe:	0	0%
		North America:	5	4.3%
		Oceania & South America:	0	0%

served that whereas the proportion of courses delivered by Computer Science departments is relatively similar, the representation of Humanities departments is much more prevalent in North America than in Europe. Likewise, multidisciplinary offers are much more represented in Europe than in North America, something which is in line with the results of ranking universities across the globe according to their multidisciplinary research, which places Europe on the top of this ranking (ranking by World University Research Rankings[8]).

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 regards to what department delivers AI ethics education can lead to very 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.

8. https://ireg-observatory.org/en/bez-kategorii/world-university-research-rankings-europe-excels-for-multidisciplinary-and-collaborative-research/

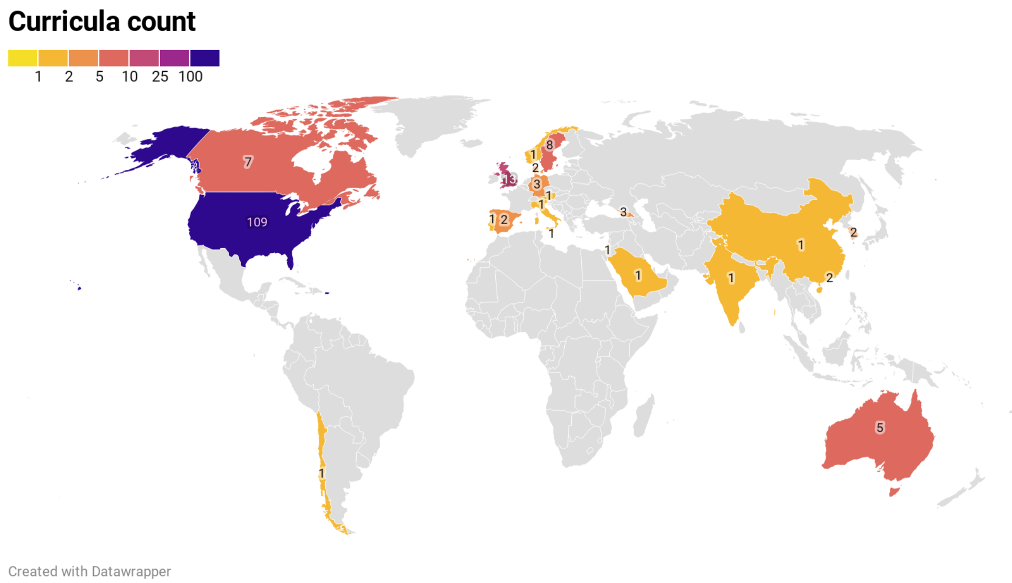

Figure 1: Geographical distribution of countries included in our dataset, with total courses count.

3.2 Trends Analysis

3.2.1 Uncovering Topics

Following the methodology described in Section 2, LDA was applied to our syllabi dataset. As a result, 10 main topics were found within our corpus. The 15 most probable words (i.e., the words with the highest probabilities), together with the semantically attached label for each uncovered latent topic, are shown in Table 2. These 10 topics can be grouped into two overarching themes: course administration (5 topics) and course content (5 topics). The intertopic distance map used to label these topics is displayed in Figure 2. The manually-assigned labels for the five course content topics are as follows: **T1** *Media & society*, **T4** *Philosophy*, **T6** *Responsibility*, **T8** *Data security*, and **T9** *Applications*.

In order to focus the analysis on the actual course content and avoid being side-tracked by the administrative content, the administrative topics were hidden in the following analysis and the remaining topics were normalized. This is possible since the probability distributions of the words identified in administrative-related and content-related topics are statistically distinct.

A relevant observation stemming from the analysis of our corpus is that ethics was not identified as a separate topic. It is also worth noting that, for every syllabus, the LDA model infers the topical decomposition, indicating what topics are found in that syllabus and in which proportion. This structure assumes that syllabi are often about several concrete topics, rather than about all topics equally.

The intertopic distance map for all topics is shown in Figure 2. This two-dimensional topic representation displays the similarity between topics with respect to their word distribution over topics; that is, the words and their corresponding probability within the topic.

Table 2: Top-10 most prevalent keywords of every topic as uncovered by the LDA, listed in descending order of their prevalence.

Topic	Label	Top-10 keywords
1	*Media & society*	medium, society, digital, optional, internet, algorithm, public, culture, platform, public
2	Adm 1	project, presentation, lecture, evaluation, read, group, reflection, team, section, report
3	Adm 2	code, point, personal, identity, review, concept, retrieve, relate, base, conduct
4	*Philosophy*	philosophy, human, moral, machine, future, robot, philosophical, read, idea, teach
5	Adm 3	program, engineering, examination, group, offer, assessment, credit, academic, module, apply
6	*Responsibility*	decision, people, principle, build, responsibility, law, explain, framework, understand, user
7	Adm 4	word, answer, text, find, short, writing, day, case, part, argument
8	*Data security*	datum, link, data, legal, law, detail, security, analysis, algorithm, bias
9	*Applications*	ai, application, video, search, develop, apply, story, support, open, table
10	Adm 5	grade, data, retrial, post, participation, email, academic, disability, resource, plagiarism

Clustered and/or overlapping nodes indicate similar word distributions, and the size of the node indicates the relative topic prevalence in the complete dataset. The topic prevalence indicates how widespread a topic is within the dataset, as all topic proportions add up to 100%. In this figure, it can be noticed that the topics **T6** *Responsibility* and **T8** *Data security* form a cluster, indicating a similar probability distribution over words (i.e., topics whose most probable words are related to some extent). The remaining course content topics, **T1** *Media & society*, **T4** *Philosophy*, and **T9** *Applications*, are isolated from each other, suggesting distinct probability distributions over words.

Media & society, the most prevalent topic in our corpus, relates to syllabi that study the relationship between technology and society. This topic includes numerous concepts related to (social) media (medium, Internet, digital, platform, community, public) and society (community, surveillance, public, practice, culture). This relates to the fact that CS and AI ethics have long been considered socio-technical systems (Jasanoff et al., 2001; Jasanoff, 2016; Harris et al., 2018).

Philosophy relates to the concepts arising from classic philosophy (philosophy, philosophical, moral, idea) with an interest over (re-)defining basic concepts (robot, human, future, AI, machine). Related syllabi often include the study of classic ethical frameworks (e.g., utilitarianism, deontology, virtue-based ethics) and fiction (movies, books). Since Humanities departments propose almost 40% of the syllabi that were included in our corpus,

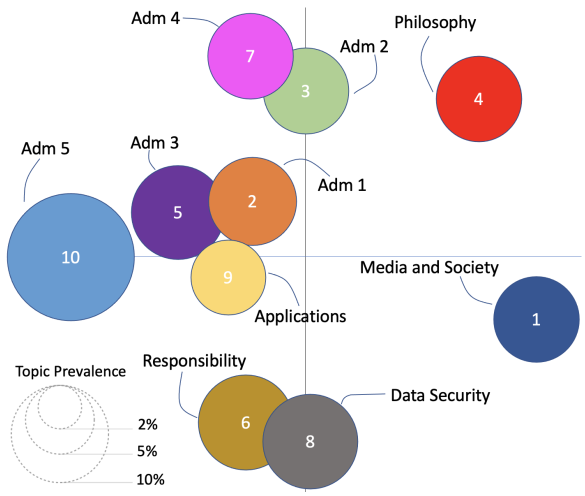

Figure 2: Intertopic distance map, showing a two-dimensional representation (via multi-dimensional scaling) of the topics. The distance between the nodes represents the topic similarity with respect to the distributions of words - closer topics tend to be correlated in the syllabi.

many of these syllabi contain more diverse aspects of Philosophy, discussing it in more detail than just ethics. Interesting keywords in this topic are robot and machine, which may indicate the philosophical concerns addressing ethical and moral issues of considering robots and machines superior to humans (Bostrom, 2014) or may indicate the philosophical concerns about machine ethics, colloquially known as "programming ethics into the machine" (Anderson & Anderson, 2011). Likewise, the occurrence of the word robots might also indicate a discussion of the (human) ethics of deploying robots for, for example, mechanization of work, eldercare, or war.[9]

Responsibility relates to the general domain of responsible design and trustworthy AI (High-Level Expert Group, 2019), as a form of an engineering practice that is sensitive to social concerns and features aspects of systems such as accountability, responsibility, and transferability (Dignum, 2019) (build, user, aspect, decision, people, law, explain, understand, impact, principle, framework, cost, responsibility).

Data security relates to regulatory and technical practices revolving around cybercrime and data management (data, datum, security, legal, law, report, analysis), including indirect regulations and concerns regarding data usage (discrimination, bias). Syllabi related to this topic tend to develop advanced legal and regulatory-related contents.

Applications relates to concrete technical matters regarding the development of AI systems (application, develop, AI, language, open, search, apply, support, table). Curricula related to this topic tend to focus on projects and specific technical matters and implementations.

Comparing our results with those from Fiesler et al. (2020), who performed a manual analysis of 115 syllabi of mainly US courses from the original set of crowd-sourced "tech-ethics" curricula, which we have also partly used, our study identified five course content

9. We thank one of the anonymous reviewers for this suggestion.

topics, compared to 15 topics in this previous study. Topic *Philosophy* is singled out by both studies. However, if grouped under themes, some of the 15 topics identified by Fiesler et al. (2020) could be related to some of our topics (e.g., Civic Responsibility & Misinformation, and Professional Ethics could relate to *Responsibility*). In addition, our study uncovered the *Applications* topic, which was not identified in the previous work by Fiesler et al. (2020). As discussed in Section 3.2.3, this can be explained by additional courses from outside the US being included in our corpus.

As a matter of 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[10], a world-leading ethical guideline. As an evaluation criterion, for every topic, we summed the number of occurrences of every of its keywords within the ACM ethics guidelines. The results are summarized in Table 3. A total of 200 occurrences of the keywords have been found in the ACM guidelines across all topics. This analysis highlights imbalances between the main approaches for teaching AI ethics and the representativeness of the taught concepts within the ACM ethics guidelines: courses closer to the *Responsibility* topic seemingly cover a conceptual background much more aligned with the concepts used by this classic ethical guidelines than courses closer to the *Applications* or *Philosophy* topics.

Table 3: Number of occurrences in the ACM ethics guidelines of the keywords of the various topics uncovered by the LDA.

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%)

3.2.2 Topic Proportion Within Syllabi

The heatmap displayed in Figure 3 shows the co-occurrence 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. For example, when setting the topic *Philosophy* as dominant, 30.4% of their content is related to the topic of *Philosophy*, leaving 69.6% of the remaining content to be spread to other topics. In this specific case, 10% of the content is related to the topic *Applications*. Note that the numbers on each row do not add up to 100% because a part of the proportion is not displayed due to being allocated to administrative topics.

Analyzing co-occurrence highlights potential relations regarding the dominant topic and the prevalence of other topics. It is interesting to note that many of these relationships are asymmetrical. Curricula with *Media & society* as a dominant topic appear to equally include all other topics, whereas syllabi dominated by topics other than *Media & society* tend to include less even distributions across topics. In particular, *Data security* and, to a

10. `https://www.acm.org/code-of-ethics`

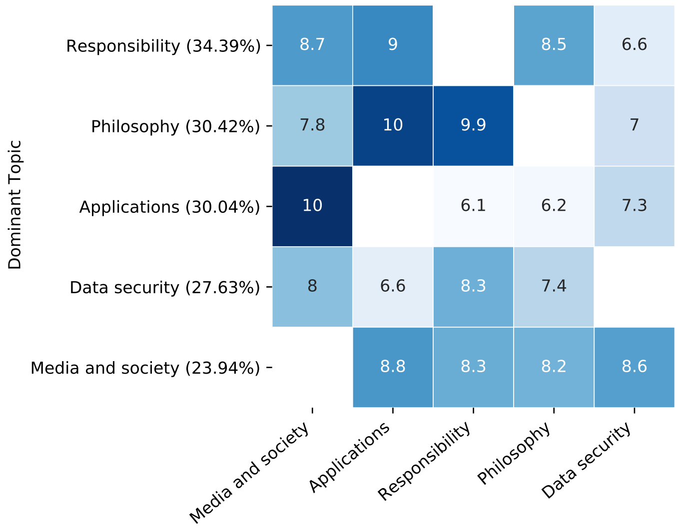

Figure 3: Heat map displaying the dominant topic (left) and the remaining average topic proportions (top) for all 166 syllabi. from the corpus - administrative topics are not included and values are normalized between 0 and 1. This indicates the extent to which documents about one main topic relate to the other uncovered topics (i.e., the degree of topic co-occurrence), decreasing from left to right. For example, documents that primarily focus on *Responsibility* also focus on *Applications* (9%) or *Media & society* (8.7%).

lesser extent, *Philosophy*, tend to be disregarded as secondary topics; whereas *Applications* and *Media & society* appear to be particularly co-occurrent secondary topics for syllabi with any dominant topic, except for, respectively, *Data security* and *Philosophy*.

Regarding more specific relations between topics, *Philosophy* and *Responsibility* appear to be mutually tied. Syllabi with *Philosophy* as the dominant topic tend to give high emphasis on *Applications*, whereas syllabi with *Applications* as the dominant topic tend to give little emphasis to *Philosophy* and *Responsibility*. Syllabi with *Data security* as the dominant topic tend to relate uniformly to other topics, with the lowest emphasis over *Applications*, this secondary topic being, however, the most prevalent for all other dominant topics. Topics with *Applications* as the dominant topic appear to give high focus to *Media & society* as secondary topic, whereas giving much less emphasis over other topics. *Applications* is the dominant topic that has the highest discrepancy in terms of co-occurrence with other topics, achieving the highest co-occurrence score with *Media & society* and the lowest co-occurrence score with *Responsibility* and *Philosophy*.

Overall, our results suggest that abstract course contents are connected with specific applications. Such applications are clearly connected with concrete needs of the society, where a mapping between the application and the provided solution can be established. These societal needs exhibit more or less evenly-distributed connections with all the other contents.

It would be interesting to consider whether other mixtures of topics would bring novel, and possibly also innovative, insight into AI ethics education. For example, syllabi where the dominant topic of *Applications* co-occurs with topics such as *Philosophy* or *Responsibility*

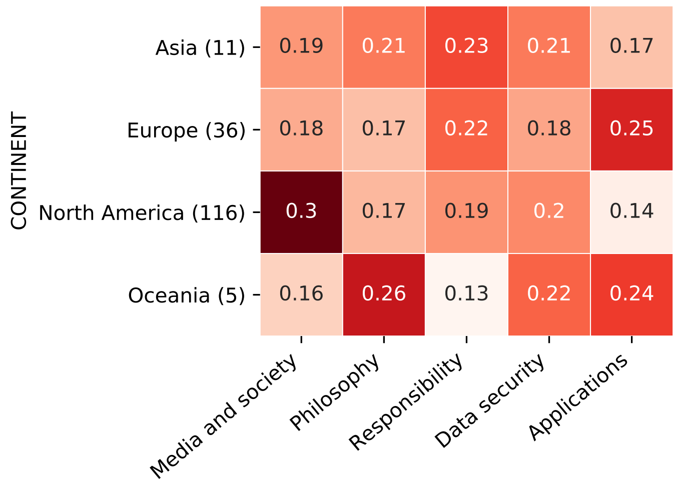

Figure 4: Distribution of the various topics depending on the continent. Each row is normalized to 1. For example, the cell (Asia, *Philosophy*) indicates that, when only considering the syllabi related to Asia, 22% of the content is from the *Philosophy* topic.

(e.g., when building a robot, besides addressing "how" to build it, students will relate the use of the robot with classic ethical frameworks and concrete methods for making it lawful, robust, and trustworthy).

3.2.3 Topics And Continents

Figure 4 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. Our results indicate that the prevalence of the five course content topics is balanced in the Asian syllabi, with a slight prevalence of the *Responsibility* topic. European syllabi tend to give more importance to the *Applications* topic. The North American syllabi shows a significantly greater emphasis on *Media & society*, while the one of Oceania to the *Philosophy* topic. The figure also indicates that the topic of *Data security* is not particularly prevalent in any of the continents, despite news on this topic being covered in the media on a daily basis across the globe. We have to note here that these results have to be read with caution, due to the limitations raised by the skewed dataset, as detailed in Section 3.4.

3.2.4 Topics and Disciplines

Figure 5 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: the curricula from computer science tend to cover a balanced mix of the various topics, with a possible greater interest given to the *Responsibility* topic, followed closely by the *Data security* and *Applications* topics; the curricula associated with multiple departments seem to give more importance to the *Philosophy* topic; the curricula associated with the humanities department tend to focus more on the *Media & society* topic;

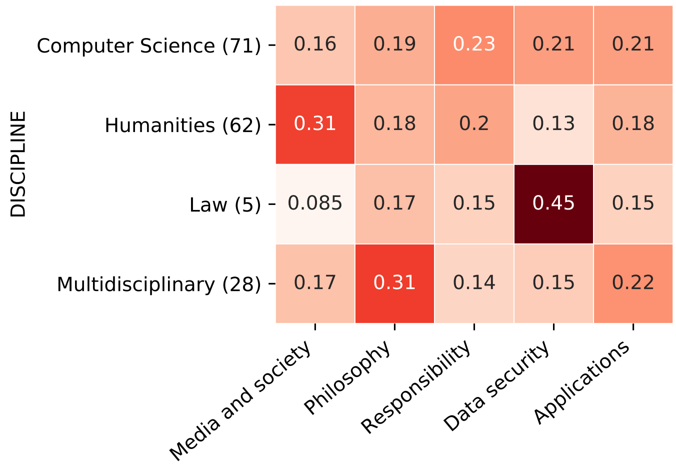

Figure 5: Distribution of the various topics depending on the discipline. Each row is normalized to 1. Number of offered courses is displayed in brackets.

the curricula associated with the law department tend to focus more on the *Data security* topic.

These results conform with stereotyped expectations of what disciplines teach what topics. However, this confirmation shows that communities are still training the learners in silos. This reinforces the critique against the fact that tunnel vision and mismatching values of what is important to learn regarding responsible design, which are still being maintained in the current curricular layout (e.g., "tech people thinking about tech activities; law people being hindrances, humanities talking about society, but being disconnected from technical realities").

3.2.5 Topics and Pedagogy

Next, in Figure 6 we answer the question: *"What is the prevalence of the various Bloom cognitive levels (Figure 6a) and teaching modalities (Figure 6b) for each given topic?"* The *Remember* Bloom cognitive level is the most prevalent for the *Data security* topic; the *Understand* Bloom cognitive level is the most prevalent for the *Media & society* topic; the *Apply* Bloom cognitive level is the most prevalent for the *Applications* topic; the *Analyze* Bloom cognitive level is not the most prevalent for any topic; the *Evaluate* Bloom cognitive level is the most prevalent for the *Philosophy* topic; and the *Create* Bloom cognitive level is the most prevalent for the *Responsibility* topic.

A per-topic analysis matches *a-priori* expectations one may have when relating a topic to a Bloom cognitive level: *Media & society* is an activity driven towards understanding society and its mechanisms within the context of AI tech development; *Philosophy* focuses on evaluating systems; *Responsibility* gives more emphasis on the responsible design of systems; *Data security* gives more emphasis on memorizing the many laws and rules that apply in various contexts, while the *Applications* topic is more centered on practical aspects.

The *Apply* teaching modalities are the most prevalent for the *Media & society* topic; the *Evaluate* teaching modalities are the most prevalent for the *Philosophy* topic; the *Analyze*

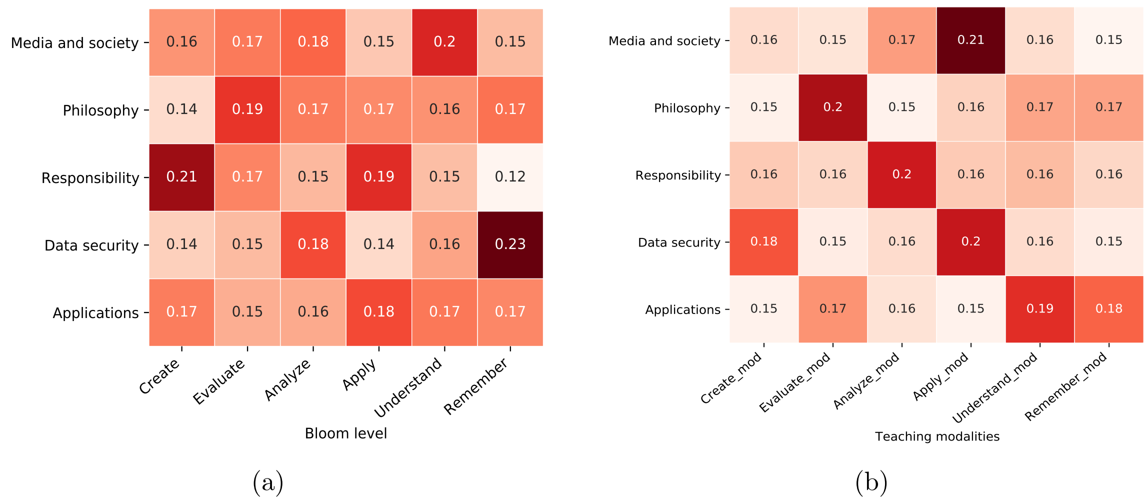

Figure 6: Each row represents the breakdown of a topic with respect to: (a) Bloom cognitive levels and (b) teaching modalities. Each row is normalized to 1.

teaching modalities are the most prevalent for the *Responsibility* topic; the *Apply* teaching modalities are the most prevalent for the *Data security* topic; the *Understand* teaching modalities are the most prevalent for the *Applications* topic.

In this case, the distribution of the prevalence of the various teaching modalities is surprising and more difficult to explain with intuitions (e.g., why is there so little emphasis on the *Apply* teaching modalities for the *Applications* topic? Why is there so much emphasis on *Create* and *Apply* teaching modalities for the *Data security* topic?). A cross-comparison between the Bloom cognitive levels and teaching modalities per topic highlights a disalignment between claimed learning objectives and pedagogical methods put in place, except for the *Philosophy* topic. For instance, syllabi with high prevalence on *Data security* tend to emphasize on developing knowledge about rules while often extensively relying on project-oriented pedagogic activities; curricula with high prevalence on *Applications* tend to rely relatively extensively on lecture and group discussions even though the score on the *Apply* Bloom cognitive level is higher than of the other ones.

A very interesting insight arises when crossing these analyses at an overall level: the topic to be taught appears to be a strongly differentiating factor with regards to the Bloom cognitive level and teaching modalities. In other words, these results match the intuition that, in current curricula, teaching different topics calls for teaching at different Bloom cognitive levels and with different teaching modalities. In turn, this suggests a topic-centered specialization of ILOs and teaching activities.

On a separate note, a cross-comparison of the average frequency of the Bloom cognitive level-related words across the syllabi highlights that these levels are used relatively balanced. This result contrasts with the one from Fiesler et al. (2020), who identified only two main categories of ILOs: recognize ethical issues and apply rules & create solutions.

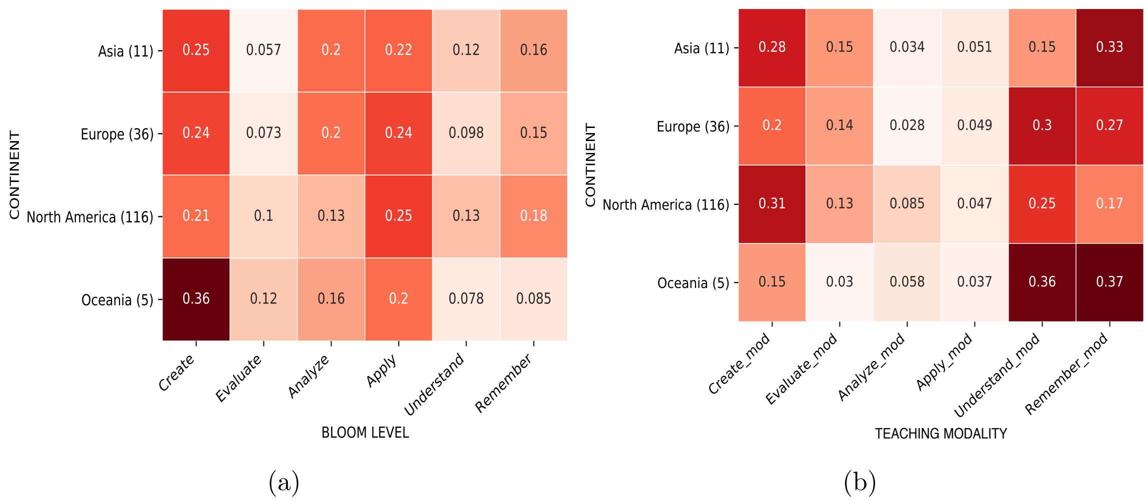

Figure 7: Breakdown of curricula from a given region into (a) Bloom cognitive levels and (b) teaching modalities. Each row is normalized to 1.

3.2.6 Pedagogy and Continents

Figure 7 addresses the question: *"What is the prevalence of the various Bloom cognitive levels (Figure 7a) and various teaching modalities (Figure 7b) in the curricula of each continent?"*. The *Create* and *Apply* Bloom cognitive levels appear to be most prevalent in all regions, with Oceania putting a stronger emphasis on the former. When it comes to teaching modalities, there are differences among continents: the *Remember* modalities are prevalent in Asia and Oceania (with the latter also emphasising the *Understand* modalities). *Understand* modalities are more prevalent in Europe and *Create* modalities are prevalent in North America. The distribution of the prevalence over the various Bloom cognitive levels and teaching modalities is relatively the same across continents, possibly indicating a common global practice.

Crossing Figure 7a and Figure 7b, it seems that there is a misalignment between ILOs and teaching modalities at continent level (e.g., there is no correspondence between *Apply* ILOs and *Apply* teaching modalities). In particular, there are recurrent misalignments across continents between Bloom cognitive level and Bloom teaching modalities: the prevalence of the *Apply* Bloom cognitive level ILOs are overrepresented with regards to the prevalence of the *Apply* teaching modalities; and the prevalence of the *Understand* Bloom cognitive level ILOs are underrepresented with regards to the prevalence of the *Understand* teaching modalities.

3.2.7 Pedagogy and Discipline

Figure 8 answers the question: *"What differentiates the syllabus from different departments in its formulation for each Bloom cognitive levels (Figure 8a) and teaching modalities (Figure 8b)?"*

One-third of the syllabus related to the *Create* and *Apply* Bloom cognitive levels overlaps with the syllabus offered by the CS department. This was not unexpected, as CS is a domain

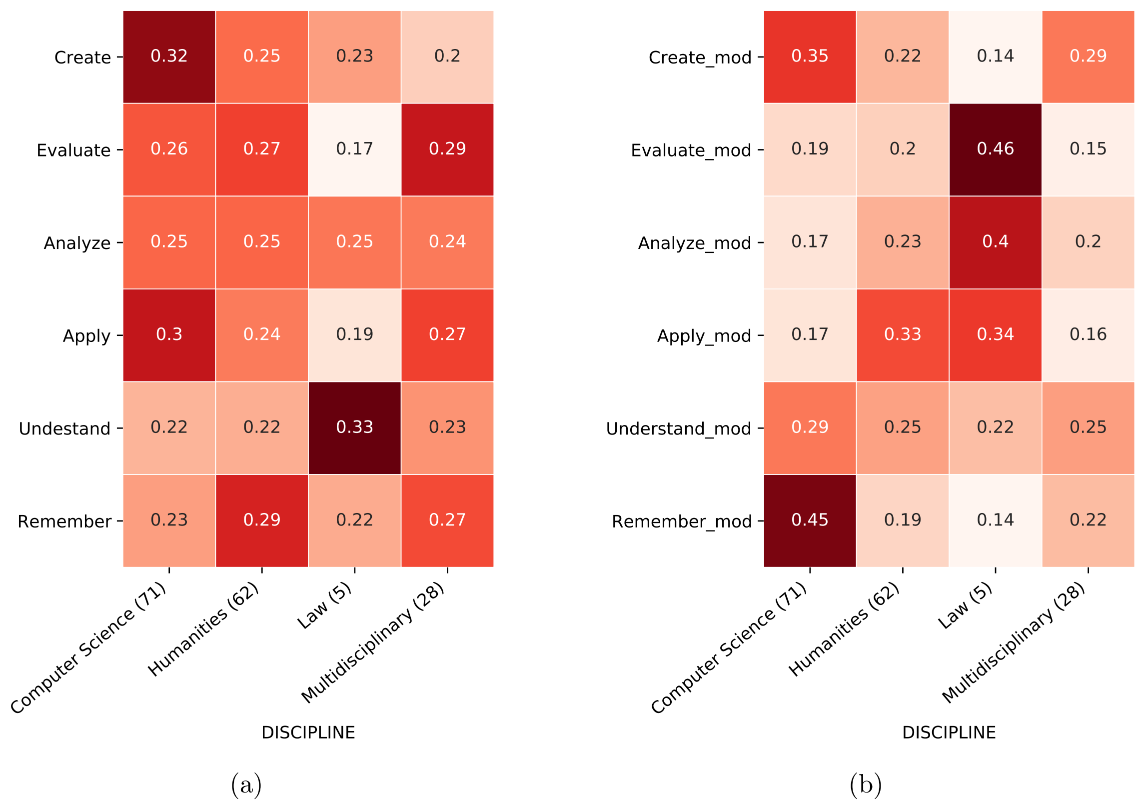

Figure 8: Composition of (a) Bloom cognitive levels and (b) teaching modalities from syllabi of courses connected to different disciplines. Each row is normalized to 1.

that proposes practical solutions. However, while this seems to be right at ILOs level, when looking at teaching modalities, there is a discrepancy, with *Remember* modalities dominating this curriculum.

One-third of the syllabi giving prevalence to the *Remember* Bloom cognitive level and *Understand* Bloom cognitive level overlaps with the curricula of the syllabi offered by the Humanities department and Law department, respectively. Nevertheless, also within these departments there is a discrepancy regarding the alignment of ILOs and teaching modalities.

At ILOs level, these results seem to confirm the stereotype that computer scientists are solution-oriented while those coming from humanities and law focus most on remembering and basic comprehension of facts, terms, and concepts.

One-third of the syllabi related to the *Evaluate* Bloom cognitive level overlaps with the syllabi offered by multiple departments. This might indicate that in such syllabi, the emphasis is put on presenting and defending opinions by making judgements based on a set of criteria, thus showing the value of syllabi taught by a multidisciplinary team in comparison with a mono-disciplinary one. However, the discrepancy between ILOs and teaching modalities described above plagues these, too.

3.3 Meta-Synthesis

A model of how AI ethics is taught across the world can be derived from a meta-synthesis of the results presented above[11]. This model 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 4. We call this model BAG: Build, Assess, and Govern. The first general approach, labelled "Build", focuses on teaching the ability to design trustworthy technical solutions. The second, labelled "Assess", focuses on teaching the ability of making judgements based on finding evidence anchored in fundamental principles of society. The third approach, labelled "Govern", focuses on teaching the core of the guiding terms and concepts needed to protect the stakeholders impacted by technical solutions.

Despite a general consensus on the fact that ethical and responsible AI design is intrinsically an interdisciplinary matter (Dignum, 2019, 2021), the BAG model highlights that the subject matter of ethics is still visibly taught in silos: specific topics, specific disciplines, specific ability levels. By teaching in silos, we are seeding the issues we are facing today, thus being in the danger of perpetuating in the future the limited practices of the present. 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 weighting 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.

Besides highlighting a problem, the BAG model also offers a potential solution for alleviating the development in silos within the limited time and resources for training AI ethics learners. This solution consists of **developing hybrid ethics courses** that blend:

11. This table and model were established based on a cross-analysis of the results from the previous sections and tables & figures. The relation between topics and disciplines is introduced in Section 3.2.3, showing that Computer Science syllabi tend to lean towards the *Responsibility* topic, Law syllabi towards the *Data security* topic, Humanities syllabi towards the *Media & society* topic, and Multidisciplinary syllabi towards the *Philosophy* topic. The relation between topics and Bloom cognitive level is identified in Section 3.2.5, notably relating the *Responsibility* topic and to a lesser degree the *Applications* topic to the *Apply* and *Create* Bloom cognitive levels; the *Philosophy* topic to the *Evaluate* Bloom cognitive level; the *Media & society* and *Data security* topics to the *Understand* and *Remember* Bloom cognitive levels. The *Applications* topic is shown to be balanced across all Bloom cognitive levels. The *Analyze* Bloom cognitive level is not favored by any particular topic. Instead of removing it from the model, we decided to associate it with the *Evaluate* Bloom cognitive level, as analysis skills are the foundation of evaluation skills. The Build, Assess, and Govern labels were associated with each of these clusters based on the general behavior and type of professional activity that can be related to the matched topics, disciplines, and Bloom cognitive level. Syllabi training computing-scientists focused on technical topics growing creative and applicative skills thematically relate to an ethical *Build*ing of AI systems. Multidiscplinary syllabi focused on philosophical and applicative topics growing analytical and evaluative skills thematically relate to an ethical *Assess*ment of AI systems. Syllabi training humanities and law students focused on knowing / understanding the issues of media, society, and data-security thematically relate to an ethical *Govern*ment of AI systems.

Table 4: 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.

	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*

1) a *generalist ethics module* taught in a multi/interdisciplinary way (e.g., including philosophy, accountability, inequality, environmental sustainability), with 2) a *domain/profession-specific ethics module* taught by a disciplinary expert (e.g., AI designer, engineering, law, medical practice, journalism). By covering in a balanced way *all of the Bloom cognitive levels* (see Figure 8a), multidisciplinary courses have the potential to develop a relatively complete array of cognitive abilities in their learner compared with courses that utilize fewer Bloom cognitive levels. This because Bloom's taxonomy is based on a specific hierarchy of learning levels, with each lower level being a vital part of moving to the next level. Such multi/interdisciplinary courses would be enriched with domain/profession-specific abilities that best match the future job of the learner. Thus, for example, data analysts taught in a teaching-and-learning system that follows our proposed solution would develop the required specific technical skills augmented with a broad spectrum of cognitive abilities allowing them to meaningfully interact with a large array of stakeholders from different domains (e.g., diverse group of users, data-protection authorities, legal departments). The teaching and learning approach proposed here enables all the involved parties to create a minimal basis for the joint understanding of the global value of ethics and the interests of the other parties.

The challenges of today and tomorrow do not come nicely arranged in disciplinary silos. 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 (e.g., *Remember* + *Apply*), or discipline bridges (e.g., 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. There are ways to abandon the BAG paradigm progressively. One could seek to promote a shift towards a shared vocabulary (Raji et al., 2021). Further action could seek a more balanced distribution of the various topics being taught in a given course. As such, as described in Section 3.2.2, courses related with *Media & society* as a dominant topic appear to be suited examples for balancing this distribution, whereas existing curricula related with *Applications* appear to be much more imbalanced. Whereas assessing whether the reason for such imbalance is beyond the scope of this paper (e.g., is it inherently more difficult to teach about philosophy in application-driven courses?), further exploration of this direction can offer means for reducing the prevalence of disciplinary silos. This being said, it must be noted that some might argue that in some cases there could be justifications for deliberately having an imbalanced distribution of topics (e.g., a strong need for specialization of the students) or that such a balance may entail sacrificing depth. Such arguments have been formulated previously in the context of the rise of multi- and interdisciplinary education (e.g., Pirrie, Hamilton, & Wilson, 1999; Benson & Miller, 1982), with significant educational and research effort put in demonstrating one perspective or the other and in providing educational tools and environments that would foster harmonized development of either approach. As expressed in Holley (2017), we believe that our proposal of a more balanced distribution of the various topics in a given course does not diminish the role of specialization in education, but rather "acknowledges that knowledge is unbounded and potential discoveries lie outside compartmentalized structures". This perspective is shared by recent scholarships regarding interdisciplinary AI ethics education (Borenstein & Howard, 2021) and several mixed-faculty courses are considered to be successful by experts[12].

3.4 Limitations of the Study

The main limitation of this study arises from the focus of analysis, that is, the course syllabi. Such texts do not necessarily faithfully represent what goes on in the classroom[13]. However, using syllabi as a proxy for the classroom activity is the best option in the case of large scale systematic assessment of what is taught in specific courses.

By relying on an LDA analysis, this study is bound by similar limitations as data-centered analysis.

First, limitations can be raised by the data collection process, as this can be incomplete (thus, it might not a representative sample) and unevenly distributed. Notably, there is an over-representation of North American syllabi and an under-representation of law syllabi. Such incompleteness and uneven distribution is partly due to the fact that Fiesler's form is open access (Fiesler, 2018) and its coverage is sensitive to whether course organizers are aware of this list and willing to contribute to it. Most probably, STEM science, technology, engineering, and mathematics (STEM) teachers working with ethics are the most likely exposed to this list and feel the most legitimate to contribute to it. As a matter of alleviating this limitation, we put considerable effort into expanding the dataset towards the inclusion of courses from under-represented regions. Despite the addition of 43 courses, the systematic inclusion of courses beyond English-speaking countries has proven to be difficult, as AI ethics curricula are not easily accessible in English or are behind Learning Management Systems only accessible internally, such as Canvas or Blackboard (this is true even for courses in English-speaking countries). Thus, the current dataset is close to the most comprehensive that could have been obtained considering this limitation.

Second, limitations can be raised by the analysis tool itself. Despite being a central pillar of automated topic modelling, LDA is not exempt from simplifications and approximations that can misguide further interpretation. The model used by the LDA is based on a statistical generative approximation of the corpus turned into 'bags of words', thus

12. These are two examples of such courses: `http://www.cs.utah.edu/~suresh/courses/ethics.html`, `https://data.berkeley.edu/data-c104-fall-2020-syllabus`. We thank one of the anonymous reviewers for these examples.
13. We thank one of the anonymous reviewers for this reflection.

losing the information about the structure of the text (e.g., negations, irony, citations). The LDA focuses on identifying discriminating keywords and best-differentiated topics, without displaying common words (e.g., ethics, in the case of our corpus) and regarding correlations between topics (Blei et al., 2003; Blei & Lafferty, 2007). Then, computing the topics and keywords relies on an approximated computing process, with a possibility of reaching a sub-optimal local optimum (i.e., another combination of keywords could have been found that is statistically more plausible or to better divide the topics).

Last, limitations can be raised by the interpretation of the results, notably, how topics were labelled and how inferences were made based on the obtained prevalence. This work has been performed manually and, as such, it involved the risk of being subjective. Following the practices established in qualitative studies (Barbour, 2001) and in LDA analysis (Baden et al., 2020), we attempted to mitigate this risk by relying on two independent researchers for establishing labels and making inferences. These activities were performed with great caution, thorough discussions, and assuming the possibility of misrepresentations arising from the data and the analysis.

4. Conclusions

Anchored in well recognized canons from pedagogy science, i.e. Bloom's taxonomy (Krathwohl, 2002) and Biggs' constructive alignment principle (Biggs & Tang, 2011), this article explores where, what, and how tech ethics are currently being taught by applying advanced statistical tools, namely a Latent Dirichlet Allocation (LDA) analysis, followed by a meta-synthesis. In contrast to previous related studies, our analysis has a wide disciplinary scope (AI ethics), a broad geographical coverage (global level), it includes courses at both undergraduate and graduate levels, it relies on automated statistical methods, and it is based on a solid pedagogy approach.

The analysis conducted in this study indicates that there are numerous significant misalignments between the Intended Learning Outcomes (ILOs) and teaching modalities put in place for training these ILOs (e.g., using lectures for training towards *Apply* Bloom cognitive level skills or using projects for *Remember*-heavy topics). Furthermore, the analysis highlights the presence of three silos related to the pedagogy strategies for teaching tech ethics, along a model which we call BAG (Build, Assess, Govern). The *Build* teaching strategy focuses on the topics responsible design, engineering and technical aspects, creative and application skills. The *Assess* learning approach, generally taught by a multidisciplinary team, focuses on developing evaluation and analysis skills through studies of philosophical topics and concrete applications. The *Govern* teaching approach, generally delivered by Humanities and Law departments, focuses on developing general knowledge and understanding of the *Data security* and *Media & society* topics.

Although we contributed to enrich the curricula dataset towards higher geographical coverage (11 courses from Asia, 33 from Europe, 116 courses from North American universities, and 6 from other regions), the dataset is still skewed towards the US. As a result, topics that are prevalent in less represented regions (e.g., Asia, Oceania), could have been overlooked by the LDA. Enriching the dataset with more diverse curricula is important in understanding how AI ethics is taught in various regions around the globe, as ethical behaviour can mean different things in different demographics. As a future work, we intend to supplement our database with courses from under-represented regions (e.g., Africa, South America, Asia) and invite fellow colleagues to help us in this endeavour.

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. As a community, such findings demonstrate an urge for concrete actions to be taken by higher-education institutions so that the engineers, lawmakers, designers, policymakers, and AI users can work together rather than against each other. It is important to challenge teaching practitioners and program coordinators, so that the usual procedures are re-evaluated. We hope that this will result in expanding the pre-established methodologies beyond the confines of traditional biases and stereotypes in teaching, in particular when it comes to what disciplines should be taught and how. This study proposes one way to address this problem, through a generic-to-specific hybrid teaching and learning approach as a possible solution for best connecting the various communities within a shared understanding of the value of AI ethics.

Acknowledgements

LV acknowledges the support of the Knut and Alice Wallenberg Foundation (project number 570080103), of the project AutograAIde "A Student-Driven Interdisciplinary Hackathon on Whether and How to Automate Grading & Assessment" (project number 570002260, The Rådet för Artificiell Intelligens, Umeå University, Sweden), and of the project "GEDAI: Growing Ethical Designers of Artificial Intelligence" at 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) and of the CRAFT Lab - Knowledge Integration and Blue Futures, UiT The Arctic University of Norway. TJ and JQ acknowledge the support of the project "Culturally-informed pro-social AI regulation and persuasion framework for Pakistan and the Muslim world", funded by Facebook Research. SG acknowledges the support of the project "digitainable" funded by the German Federal Ministry for Education and Research (BMBF). RV acknowledges the financial support by Swedish Research Council (VR).

Appendix A. Search Strategy for Expanding the Initial Dataset

In order to extend the dataset, we used various search terms that were related to AI Ethics education. In order to ensure Google gives good results irrespective of the region we are searching from a VPN was used to simulate searches from different regions. A full list of search terms along with the regions and VPNs used is included in Table 5. Furthermore, when considering a course for inclusion within our dataset the contents of the syllabi were checked to make sure it is in line with our aims for this study. As a result we manually ensured that terms related to AI, like data science, machine learning, artificial intelligence, deep learning etc., were among the central themes of the curricula.

Table 5: Search terms that were used to identify courses from across the globe.

Term	Search Provider	VPN Location
AI Ethics Curricula	Google	
Ethics Syllaby	Google	
AI Ethics Curricula	Bing	
Ethics Syllaby	Bing	
AI Ethics Curricula Asia	Google	
AI Ethics Curricula Asia	Google	Turkey
AI Ethics Curricula China	Google	
AI Ethics Curricula China	Bing	
AI Ethics China	Google	
AI Ethics Asia	Google	
AI Ethics Pakistan	Google	
AI Ethics India	Google	India
AI Ethics Syllabus Australia	Google	
AI Ethics Syllabus Africa	Google	
AI Ethics Syllabus China	Google	
AI Ethics Syllabus Africa	Google	South Africa
AI Ethics Syllabus Europe	Google	Germany
Ethics Syllabus Europe	Google	
Ethics Syllabus	Google	France
Machine Ethics Curriculum	Google	England
AI Ethics Russia	Google	Russia
Courses on AI Ethics	Google	South Africa
Courses on AI Ethics	Google	Germany
Courses on AI Ethics	Google	England
Curricula AI Ethics	Google	Saudia
Ethics Curricula Middle East	Google	Turkey
Computer Ethics Curricula	Google	
Artificial Intelligence Curricula Ethics	Google	
Artificial Intelligence Ethics syllabus China	Google	
Artificial Intelligence Ethics syllabus India	Google	
Artificial Intelligence Ethics syllabus Australia	Google	
Artificial Intelligence Ethics syllabus Sweden	Google	
Artificial Intelligence Ethics syllabus Brazil	Google	
Artificial Intelligence Ethics syllabus Canada	Google	
Artificial Intelligence Ethics syllabus Germany	Google	
Artificial Intelligence Ethics syllabus Pakistan	Google	
Artificial Intelligence philosophy syllabus	Google	
Ethics and AI Technology syllabus	Google	
Computer Science Ethics syllabus	Google	
AI Philosophy syllabus	Google	
Artificial Intelligence Governance syllabus	Google	

Appendix B. Stopwords

We used the nltk's list of English stopwords to make sure our topics were not influenced by these redundant words. These stopwords are listed below.

> "i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "you're", "you've", "you'll", "you'd", "your", "yours", "yourself", "yourselves", "he", "him", "his", "himself", "she", "she's", "her", "hers", "herself", "it", "it's", "its", "itself", "they", "them", "their", "theirs", "themselves", "what", "which", "who", "whom", "this", "that", "that'll", "these", "those", "am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had", "having", "do", "does", "did", "doing", "a", "an", "the", "and", "but", "if", "or", "because", "as", "until", "while", "of", "at", "by", "for", "with", "about", "against", "between", "into", "through", "during", "before", "after", "above", "below", "to", "from", "up", "down", "in", "out", "on", "off", "over", "under", "again", "further", "then", "once", "here", "there", "when", "where", "why", "how", "all", "any", "both", "each", "few", "more", "most", "other", "some", "such", "no", "nor", "not", "only", "own", "same", "so", "than", "too", "very", "s", "t", "can", "will", "just", "don", "don't", "should", "should've", "now", "d", "ll", "m", "o", "re", "ve", "y", "ain", "aren", "aren't", "couldn", "couldn't", "didn", "didn't", "doesn", "doesn't", "hadn", "hadn't", "hasn", "hasn't", "haven", "haven't", "isn", "isn't", "ma", "mightn", "mightn't", "mustn", "mustn't", "needn", "needn't", "shan", "shan't", "shouldn", "shouldn't", "wasn", "wasn't", "weren", "weren't", "won", "won't", "wouldn", "wouldn't"

Appendix C. Bloom's Taxonomy Verbs

Table 6: Bloom's taxonomy verbs used in our study (taken from University of Arkansas website (https://tips.uark.edu/blooms-taxonomy-verb-chart/)).

Remember	*Understand*	*Apply*	*Analyze*	*Evaluate*	*Create*
Cite	Add	Acquire	Analyze	Appraise	Abstract
Define	Approximate	Adapt	Audit	Assess	Animate
Describe	Articulate	Allocate	Blueprint	Compare	Arrange
Draw	Associate	Alphabetize	Breadboard	Conclude	Assemble
Enumerate	Characterize	Apply	Break down	Contrast	Budget
Identify	Clarify	Ascertain	Characterize	Counsel	Categorize
Index	Classify	Assign	Classify	Criticize	Code
Indicate	Compare	Attain	Compare	Critique	Combine
Label	Compute	Avoid	Confirm	Defend	Compile
List	Contrast	Back up	Contrast	Determine	Compose
Match	Convert	Calculate	Correlate	Discriminate	Construct
Meet	Defend	Capture	Detect	Estimate	Cope
Name	Describe	Change	Diagnose	Evaluate	Correspond
Outline	Detail	Classify	Diagram	Explain	Create

Point	Differentiate	Complete	Differentiate	Grade	Cultivate
Quote	Discuss	Compute	Discriminate	Hire	Debug
Read	Distinguish	Construct	Dissect	Interpret	Depict
Recall	Elaborate	Customize	Distinguish	Judge	Design
Recite	Estimate	Demonstrate	Document	Justify	Develop
Recognize	Example	Depreciate	Ensure	Measure	Devise
Record	Explain	Derive	Examine	Predict	Dictate
Repeat	Express	Determine	Explain	Prescribe	Enhance
Reproduce	Extend	Diminish	Explore	Rank	Explain
Review	Extrapolate	Discover	Figure out	Rate	Facilitate
Select	Factor	Draw	File	Recommend	Format
State	Generalize	Employ	Group	Release	Formulate
Study	Give	Examine	Identify	Select	Generalize
Tabulate	Infer	Exercise	Illustrate	Summarize	Generate
Trace	Interact	Explore	Infer	Support	Handle
Write	Interpolate	Expose	Interrupt	Test	Import
	Interpret	Express	Inventory	Validate	Improve
	Observe	Factor	Investigate	Verify	Incorporate
	Paraphrase	Figure	Layout		Integrate
	Picture graphically	Graph	Manage		Interface
	Predict	Handle	Maximize		Join
	Review	Illustrate	Minimize		Lecture
	Rewrite	Interconvert	Optimize		Model
	Subtract	Investigate	Order		Modify
	Summarize	Manipulate	Outline		Network
	Translate	Modify	Point out		Organize
	Visualize	Operate	Prioritize		Outline
		Personalize	Proofread		Overhaul
		Plot	Query		Plan
		Practice	Relate		Portray
		Predict	Select		Prepare
		Prepare	Separate		Prescribe
		Price	Subdivide		Produce
		Process	Train		Program
		Produce	Transform		Rearrange
		Project			Reconstruct
		Provide			Relate
		Relate			Reorganize
		Round off			Revise
		Sequence			Rewrite
		Show			Specify
		Simulate			Summarize
		Sketch			

		Solve			
		Subscribe			
		Tabulate			
		Transcribe			
		Translate			
		Use			

References

Abad, C. L., Ortiz-Holguin, E., & Boza, E. F. (2021). Have we reached consensus? an analysis of distributed systems syllabi. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pp. 1082–1088.

Ahmad, F. (2014). Computer science & engineering curricula and ethical development. In *2014 International Conference on Teaching and Learning in Computing and Engineering*, pp. 220–225. IEEE.

Anderson, M., & Anderson, S. L. (2011). *Machine ethics*. Cambridge University Press.

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

Baden, C., Kligler-Vilenchik, N., & Yarchi, M. (2020). 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.

Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: a case of the tail wagging the dog?. *Bmj*, *322*(7294), 1115–1117.

Becker, B. A., & Fitzpatrick, T. (2019). What do cs1 syllabi reveal about our expectations of introductory programming students?. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, pp. 1011–1017.

Benson, T. L., & Miller, R. C. (1982). Five arguments against interdisciplinary studies. *Issues in Interdisciplinary Studies*.

Bielefeldt, A. R., Polmear, M., Knight, D., Canney, N., & Swan, C. (2019). 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.

Biggs, J. B., & Tang, C. (2011). *Teaching for quality learning at university*. McGraw-hill education (UK).

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The annals of applied statistics*, *1*(1), 17–35.

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bloom, B., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1965). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. David McKay Company.

Borenstein, J., & Howard, A. (2021). Emerging challenges in ai and the need for ai ethics education. *AI and Ethics*, *1*(1), 61–65.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Burton, E., Goldsmith, J., & Mattei, N. (2018). How to teach computer ethics through science fiction. *Communications of the ACM*, *6*, 54–64.

Chand, D. R. (1974). Computer science education in business schools. In *ACM Special Interest Group on Computer Science Education Bulletin*, Vol. 6, pp. 91–97.

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Computers & Education*, *151*, 103855.

Debortoli, S., Müller, O., Junglas, I., & vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems*, *39*(1), 7.

Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education*, *19*, 1–11.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer Nature Switzerland.

DiMaggio, P., Nag, M., & Blei, D. (2013). 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.

Fiesler, C. (2018). Tech Ethics Curricula: A Collection of Syllabi. `https://cfiesler.medium.com/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18`.

Fiesler, C., Garrett, N., & Beard, N. (2020). What do we teach when we teach tech ethics? a syllabi analysis. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pp. 289–295.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

Fuller, U., Johnson, C. G., Ahoniemi, T., Cukierman, D., Hernán-Losada, I., Jackova, J., Lahtinen, E., Lewis, T. L., Thompson, D. M., Riedesel, C., & Others (2007). Developing a computer science-specific learning taxonomy. *ACM SIGCSE Bulletin*, *39*(4), 152–170.

Garrett, N., Beard, N., & Fiesler, C. (2020). 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)*, pp. 272–278.

Gehman, W. G. (1984). Responsible Software Engineering. In *ACM Computer Science Conference*, p. 177.

Gitelman, L. (2013). *"Raw Data" Is an Oxymoron*. The MIT Press.

Goldsmith, J., & Burton, E. (2017). Why teaching ethics to AI practitioners is important. In *Proceedings of the... AAAI Conference on Artificial Intelligence*.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), 267–297.

Grosz, B. J., Grant, D. G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., & Waldo, J. (2019). Embedded ethics: integrating ethics across cs education. *Communications of the ACM*, *62*(8), 54–61.

Gupta, S., Motlagh, M., & Rhyner, J. (2020). The digitalization sustainability matrix: A participatory research tool for investigating digitainability. *Sustainability*, *12*(21), 9283.

Harris, C., Pritchard, M., Rabins, M., James, R., & Englehardt, E. (2018). *Engineering Ethics: Concepts and Cases*. Cengage Learning.

Henriksson, H., & Grunewald, E. W. (2020). The next sustainability frontier is digital. In *Sustainability Leadership*, pp. 223–244. Springer.

High-Level Expert Group (2019). Ethics guidelines for trustworthy AI. Tech. rep., European Commission.

Holley, K. (2017). Interdisciplinary curriculum and learning in higher education. In *Oxford research encyclopedia of education*.

Homkes, R., & Strikwerda, R. A. (2009). 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*, pp. 133–137.

Hughes, J., Plaut, E., Wang, F., von Briesen, E., Brown, C., Cross, G., Kumar, V., & Myers, P. (2020). Global and local agendas of computing ethics education. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, pp. 239–245.

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001.

Jang, K.-a., Park, M., & Kim, W.-J. (2018). Topic analysis for the fourth industrial revolution using LDA and centrality analysis in Korea. In *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, pp. 103–107. IEEE.

Jasanoff, S. (2016). *The Ethics of Invention: Technology and the Human Future*. W. W. Norton.

Jasanoff, S., Markle, G., Peterson, J., & Pinch, T. (2001). *Handbook of Science and Technology Studies*. SAGE Publications.

Jean, N., Burke, M., Xie, M., Davis, M. W., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*, 790–794.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgement*. Little, Brown Spark.

Khademi, K., & Hui, B. (2020). Towards understanding the HCI education landscape. In *Koli Calling'20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*, pp. 1–2.

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

Lin, C.-H., Yu, C.-C., Shih, P.-K., & Wu, L. Y. (2021). Stem based artificial intelligence learning in general education for non-engineering undergraduate students. *Educational Technology & Society*, *24*(3), 224–237.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., & Others (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2-3), 93–118.

Marçal, I., Garcia, R. E., Eler, D., & Correia, R. C. M. (2020). A strategy to enhance computer science teaching material using topic modelling: Towards overcoming the gap between college and workplace skills. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pp. 366–371.

Markham, A., & Buchanan, E. (2012). Ethical decision-making and internet research: Version 2.0. recommendations from the AoIR ethics working committee. `aoir.org/reports/ethics2.pdf`.

Martin, C. D., Huff, C., Gotterbarn, D., & Miller, K. (1996). A framework for implementing and teaching the social and ethical impact of computing. *Education and Information Technologies*, *1*(2), 101–122.

Matsuda, Y., Sekiya, T., & Yamaguchi, K. (2018). Curriculum analysis of computer science departments by simplified, supervised LDA. *Journal of Information Processing*, *26*, 497–508.

Mitchell, T. M. (1980). The need for biases in learning generalizations. Tech. rep., Computer Science Department, Rutgers University, CMB-TR-117.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, *33*(4), 659–684.

Moller, F., & Crick, T. (2018). A university-based model for supporting computer science curriculum reform. *J. Comput. Educ.*, *5*, 415–434.

Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2021). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. *Journal of Business Research*, *124*, 389–404.

Nasir, O., Muntaha, S., Javed, R. T., & Qadir, J. (2021). Work in progress: Pedagogy of engineering ethics: A bibliometric and curricular analysis. In *2021 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1553–1557. IEEE.

Nielsen, N. R. (1972). Social responsibility and computer education. In *Proceedings of the ACM SIGCSE Technical Symposium on Computer Science Education*, pp. 90–96.

O'Neil, C. (2017). The ivory tower can't keep ignoring tech. *The New York Times*, *14*.

Pirrie, A., Hamilton, S., & Wilson, V. (1999). Multidisciplinary education: some issues and concerns. *Educational Research*, *41*(3), 301–314.

Qadir, J., Shafi, A., Al-Fuqaha, A., Taha, A.-E. M., Yau, K. L. A., Ponciano, J., Hussain, S., Imran, M. A., Muhammad, S. S., Rais, R. N. B., Rashid, M., & Tan, B. L. (2020). Outcome-based engineering education: A global report of international obe accreditation and assessment practices. *ASEE Annual Conference Proceedings*.

Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021). You can't sit with us: Exclusionary pedagogy in AI ethics education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 515–525.

Rusch, T., Hofmarcher, P., Hatzinger, R., Hornik, K., & Others (2013). Model trees with topic model preprocessing: An approach for data journalism illustrated with the wikileaks afghanistan war logs. *Annals of Applied Statistics*, *7*(2), 613–639.

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., Dewar, N., & Beard, N. (2019). Integrating ethics within machine-learning courses. *ACM Transactions on Computing Education*, *19*, 1–26.

Sekiya, T., Matsuda, Y., & Yamaguchi, K. (2010). Analysis of computer science related curriculum on LDA and Isomap. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*, pp. 48–52.

Soden, R., Skirpan, M., Fiesler, C., Ashktorab, Z., Baumer, E. P., Blythe, M., & Jones, J. (2019). Chi4evil: Creative speculation on the negative impacts of hci research. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–8.

Syed, S., Borit, M., & Spruit, M. (2018). Narrow lenses for capturing the complexity of fisheries: A topic analysis of fisheries science from 1990 to 2016. *Fish and Fisheries*, *19*(4), 643–661.

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

Tzachor, A., Whittlestone, J., Sundaram, L., & Heigeartaigh, S. O. (2020). Artificial intelligence in a crisis needs ethics with urgency. *Nature Machine Intelligence*, *2*(7), 365–366.

van Berkel, N., Papachristos, E., Giachanou, A., Hosio, S., & Skov, M. B. (2020). A systematic assessment of national artificial intelligence policies: Perspectives from the Nordics and beyond. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pp. 1–12.

Vinuesa, R., & Sirmacek, B. (2021). Interpretable deep-learning models to help achieve the Sustainable Development Goals. *Nature Machine Intelligence*, *3*, 926.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, *11*(1), 1–10.

Wong, M. K., Wu, J., Ong, Z. Y., Goh, J. L., Cheong, C. W. S., Tay, K. T., Tan, L. H. S., & Krishna, L. K. R. (2019). Teaching ethics in medical schools: A systematic review from 2000 to 2018. *Journal of Medical Education*, *18*, 226–250.

Yang, S. J., Ogata, H., Matsui, T., & Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, *2*, 100008.