Embedding Responsible AI in Technical Education Curriculum: A Case Study in an Asynchronous Online Advanced Data Analytics Course

Rhodora Abadia [1]

[1] *University of South Australia, Adelaide, South Australia, Australia*

Abstract

The rapid growth and widespread adoption of Artificial Intelligence (AI) highlighted the urgent need for higher education institutions to reform how ethics in computing, data science, and related fields are taught. While computing and professional ethics are typically included in the curriculum, they are often presented from a social science, legal, or philosophical perspective. This paper presents a case study on integrating responsible AI principles into an asynchronous online advanced data analytics course from a technical perspective. The study found that embedding ethics throughout the curriculum using experiential learning and applied ethics was effective in fostering students' comprehension and application of responsible AI concepts. However, challenges included finding relevant case scenarios, lack of expertise among teachers, and developing suitable activities and assessments. Best practices identified include utilizing real-world case studies, implementing hands-on ethical coding exercises, and adopting interdisciplinary approaches. Lessons learned emphasize the importance of timing, practical application, and flexible curriculum design. This approach enabled students to assess and plan for the human consequences of AI applications, and design and implement risk mitigation strategies. The study represents a step forward in preparing students for ethical challenges in AI, while highlighting areas for future work, including teacher training and curriculum-wide integration of responsible AI principles.

Keywords

Responsible AI in Education, Ethics in Computing Education, Principles of Responsible AI

1. Introduction

Teaching ethics in computing-related degrees are often required by accrediting bodies or board with oversight of the profession. International standards such as ABET's accreditation standard [1] and ACM's Code of Ethics [2], and Australia's ACS Code of Ethics [3] lay the foundation of ethics education here in Australia. The increasing permeation of Artificial Intelligence (AI) systems lead to the concerns in the use of data, lack of trust and challenges with the use of AI. Governments and major players in the field have all started specifically looking at their practices and evaluate how they can scale up their use of AI systems and at the same time minimize AI risks. Every AI, data or related professionals has a responsibility to understand the social, political, and ethical consequences of their work. The higher education sector who produces these specialists who are more likely to use AI algorithms are compelled to rethink how they can teach the responsible use of AI. A survey made by [25] looked at machine learning courses and found that students were not taught ethics and if they were, students enroll in a stand-alone ethics elective course. Most approaches to ethics education are case-based teaching where students are presented with cases, and they respond by discussing their approach and decisions using existing ethical frameworks [6] [7][30]. Although this approach seems to be successful for some professions, education in Artificial Intelligence-related courses struggle to provide sound training in ethics that are essential for them to be successful in their career. Frauenberger, Rauhala, & Fitzpatrick [16] argued that ethical concerns are still managed in the mindset of past paradigms that largely remain static and have determined outcomes. It is only in recent years that ethical topics are integrated and infused in computing related curricula [15] [18].

7th Conference on Technology Ethics (TETHICS2024), November 6–7, 2024, Tampere, Finland
rhoda.abadia@unisa.edu.au
0000-0002-8265-0503

This paper presents the experiences in integrating the responsible use of AI into an upper-level undergraduate data analytics course. Specifically, the following are the objectives of this study:

1. Describe how the course was designed to embed the principles of responsible AI in an asynchronous online course.
2. Discuss challenges faced in the course design.
3. Present best practices and lessons learned that can be used to replicate the success of the case.

The technical practical (applied) ethics approach was employed to address ethical issues in AI education. An investigation of key ethical concerns and their application within the technical practical activities proved conducive to ethical reasoning, as these activities were specifically designed to align with the subject matter of the ethical inquiry. The curriculum emphasized topics highlighted by the ACM Conference on Fairness, Accountability and Transparency (ACM FAccT). Moreover, the Principles of Responsible AI, as established by industry leaders such as Microsoft [26], Google [4], IBM [27], and Meta [13] were adapted in for the course. The UNESCO recommendation on Ethics of Artificial Intelligence [45] and European Commission ethics guideline for trustworthy AI [46], who were also used as a guide in this course, emphasize similar core principles such as fairness, transparency, privacy, safety, and human oversight. They also stress the importance of accountability and the need for AI to benefit society.

2. Theoretical Background

This section presents the relevant literatures, theories and frameworks that were used in the development of the course.

2.1. Principles of Responsible AI

The term responsible AI encompasses a variety of technical, legal, and ethical considerations that apply not only to AI but also to data analytics and data science. The applications of AI have grown exponentially throughout the years, and current laws, policies, and standards have not caught up with the unique challenges and risks that AI poses and the changing ways society is adopting these technologies. This is why there is a need for comprehensive guidelines. While there is increasing pressure on organizations to scale up their use of AI, there is also a growing demand for the responsible use of AI.

Several big organizations and international bodies have identified principles that guide the development and use of AI applications. UNESCO's Recommendation on the Ethics of Artificial Intelligence [45] outlines ten key principles: Proportionality and Do No Harm, Safety and Security, Fairness and Non-Discrimination, Sustainability, Right to Privacy and Data Protection, Human Oversight and Determination, Transparency and Explainability, Responsibility and Accountability, Awareness and Literacy, and Multi-stakeholder and Adaptive Governance. Similarly, the European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence [246 propose seven key requirements: Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination and Fairness, Societal and Environmental Well-being, and Accountability. These frameworks demonstrate a growing consensus on the fundamental principles of responsible AI development and deployment.

Major tech companies have also developed their own frameworks. Microsoft defines responsible AI as an "advancement of AI driven by ethical principles that put people first; and making sure AI systems are developed responsibly and in ways that warrant people's trust" [26]. Accenture defines responsible AI as "the practice of designing, developing, and deploying AI with good intention to empower employees and businesses, and fairly impact customers and society" [28]. Google AI [4], IBM [27], Microsoft [26], and Meta [13] have common principles that can be classified into five categories: fairness and inclusion, transparency and explainability, reliability and safety, privacy and security, and accountability.

Fairness refers to the lack of bias. Including and involving diverse people creates data diversity which helps prevent bias. Transparency and explainability refer to the need for AI systems and their related components to be understandable, explainable, and interpretable. Reliability and safety refer to AI systems being built and tested for safety, performing what they were originally designed to do, responding safely to new situations, and resisting unintended manipulation. Privacy and security principles include the incorporation of privacy and security in the design of AI systems. The goal is to ensure that all data, whether personal and/or sensitive, should be used ethically in all AI systems. Lastly, the accountability principle refers to making people who design, develop, and deploy AI systems accountable for the impact of actions and decisions of the technology. This includes making users of the system accountable for how they use the systems.

A recent big player in the development of artificial general intelligence (AGI) applications is OpenAI. OpenAI's principles, however, do not seem to address most of the concerns raised in the responsible use of AI [29]. OpenAI's principles focus on broadly distributed benefits, long-term safety, technical leadership, and cooperative orientation.

While there are some variations in emphasis and terminology, these various frameworks and guidelines share common themes: algorithmic fairness and diversity, reliability and safety, privacy and security, and transparency and accountability. Thes principles form the foundation for the ethical development and deployment of AI systems across various domains and applications.

2.2. Teaching Ethics

Teaching ethics can be classified as theoretical and applied ethics to distinguish where is the focus of the ethical investigations [16]. Theoretical ethics often concerns itself with the understanding of the nature, language, and reasoning in ethics. Applied ethics is the practical application of the ethical theory to a problem specific to the field in question.

There are several best practices in teaching ethics in law, health, engineering, Information Technology, and business. Azim & Shamim [5] looked at educational theories that inform the education strategies for teaching ethics in undergraduate medical education and found that reflection, constructivist, and experiential learning theories are best suited in guiding strategies in teaching ethics. A European survey that examined how computer ethics are taught in computer science, or related degrees found that 66 % of the universities teach ethics as part of their computing degree [33].

The case method seems to be the common approach particularly in the field of computing, science, and engineering [6] [17] [32]. There are a variety of cases used in the case method, some are used independent of the others while some can be a combination of the others. Cases can be classified as narrative vs dialogue, single perspective vs multi perspective, hypothetical vs actual, stories vs problems, view as reader a participant vs an outside judge; success vs positive, single issue vs multi-issue, single stave vs multi-stage, ordinary vs technical language, personal vs policy, living case vs published cases [11]. All of these can be used in creating case studies for teaching ethics, it is highly suggested that technical courses use technical language where students have the same technical training or background as the instructor. Stratton [34] emphasized that moral judgement, as a skill, must be practiced and simulations provide the students-controlled environment.

Stavrakakis etal [32] and Lewis & Stoyanovich [40] argued that traditional ethics education for computing and/or data related curriculum are often taught separately as an ethics course and may not include practical and timely training on how to weigh the consequences that can be applied in their profession. Skirpan et al. [32] use of ethical thinking throughout the process of learning the fundamentals of human-centered design while Lewis & Stoyanovich[40] used algorithmic development to teach ethical data science. Both studies have shown increased interests among students. In their study of incorporating social issues of computing in a liberal art setting, Davis & Walker [12] identified ways social issues can be addressed in the technical topics and summative assessments. The challenges they identified is that framing an exam question related to social issues can seem awkward and students raise issues that have been raised in other courses. Lack of staff availability and expertise was cited as a reason for not teaching computing ethics [33] while Wueste [37] argued that curricular time demands in science and engineering disciplines are major obstacles in successfully integrating ethics and calls a need for professional development for teachers.

In traditional ethics education for computing, the case method approach, as demonstrated in related studies, is often presented as a distinct course or topic within a curriculum, lacking practical training on reflective application in a professional context. This study aims to bridge this gap by integrating case studies, practical applications, and ethical considerations into the technical content of the course. By intertwining technical skills and knowledge acquisition with a heightened awareness of ethical implications, students are provided with a comprehensive learning experience. The objective is to enhance ethics education by offering a pedagogical approach that effectively delivers the learning outcomes encompassing both technical proficiency and responsible AI applications.

2.3. Asynchronous Online Course Design

Asynchronous online is a kind of online learning where students are allowed to study at their own time. The communication of teaching and learning does not happen at the same [38]. Although studies have shown that a well-designed course can increase student satisfaction and their learning experiences, there is no consensus on what the guidelines of a good online course design are [42]. A common approach is to use an online course design template or a checklist to provide consistency to students in both accessing and navigating the course site, and assists teachers in saving time, reducing cognitive load, and meet compliance requirements [21][43]. All course content design require a systematic approach whether taught online or face-to-face. Several studies report that a systematic approach in course content design and they focus on alginment of course learning outcomes, activities and assessments [39][41] [42]. Having a a strong online course objectives and the selection of teaching methods to achieve them is important in an online course design. Wankel [36] suggested that the course objectives is the key to a successfully online course design and that the design process can be structured to four essential activities: sharing of information, illustrate skills, guide practice of skills and ensure that learning occurred.

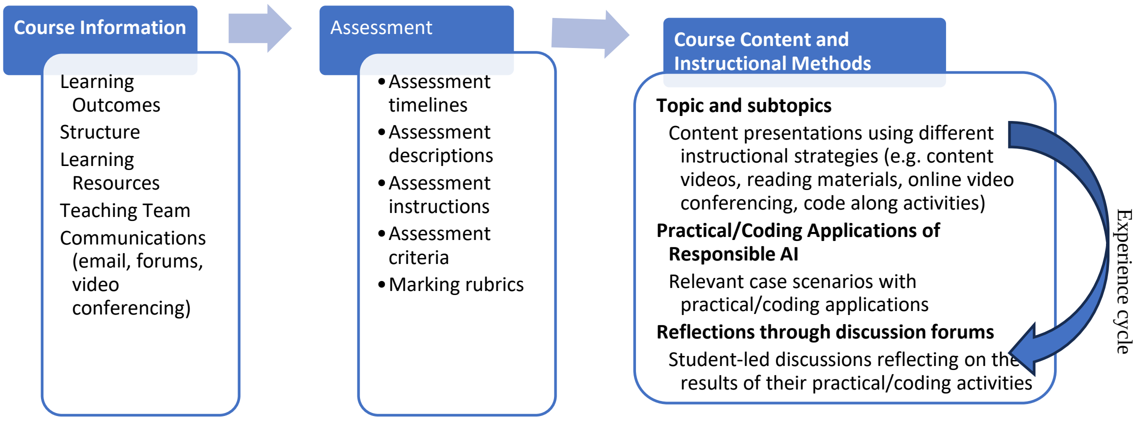
When sharing information (deliver contents), teachers have variety of choices and are not only limited to text-based approach. They can curate existing digital resources, use audio/video recordings, conferencing systems, web-based and learning management system tools or even immersive technologies (e.g., virtual reality). One strategy employed to increase accessibility [23] and teacher presence is the use of videos. There are several factors that were considered in the design of videos including how information should be presented and how it can be supported by providing additional learning activities [44]. These activities may involved additional tasks that a student may peform to reinforce what has been learned in the videos. Of all the different learning tasks, authentic activities in online learning has shown many opportunities to increase learning. Authentic activites involve presenting students with complex and extended cae scenarios that allow them to fully engage in problem-solving within realistic situations that closely resemble the context where the knowledge they are acquiring can be practically applied [19].

3. Embedding Principles of Responsible AI in Asynchronous Online Course

Working on this theoretical background, a course was developed that integrates case studies, practical applications, and ethical considerations into the technical content of the course. The common principles of responsible AI were identified. Among the different principles, Microsoft's grouping of the Principles of Responsible AI [26] were adapted as it covers all the principles identified by different AI organizations. The principles were classified as algorithmic fairness and diversity, reliability and safety, privacy and security, and transparency and accountability.

The technical practical (applied) ethics approach in teaching ethical issues in AI was used. The course was designed such that the principles of responsible AI are embedded into experiential learning. Experiential learning theory as synthesized by Kolb [22] defines learning as "a process where knowledge is created through the transformation of experience". The experience cycles of discussion, feedback, and practice and application in real-life context helps students apply and connect theoretical knowledge with real-life applications. This experience cycle is repeated throughout the course and was used as the basis of the design of this course. Figure 1 shows the framework that was used in designing the online course activities.

Figure 1: Integrating Technical (Applied) Responsible AI in the Online Course Design



The course information contains an overview of the course structure, requirements, learning outcomes, and resources. A separate section is dedicated the assessments in the course. One of the learning outcomes specifically states the knowledge, competencies and skills that are expected of the student to acquire related to Responsible AI. The assessment methods are mapped to these learning outcomes. The course is divided into unit topics. In the case study, the topics are divided into weekly topics. The cycle is students learn about the topic followed by practical activities that are related to the topic. The practical activity often includes coding activities where students apply the principles of responsible AI in the case scenarios using coding. This activity is often followed by discussions where students share their reflection on the case scenarios.

3.1. Case Study: Teaching Responsible AI in an Advanced Data Analytics Course

The course is an undergraduate-level advanced courses where data analytics students are introduced to AI topics such as reinforcement learning, computer vision and natural language processing. Students who are enrolled in this course have the basic skills and knowledge in machine learning, specifically, artificial neural networks and intermediate programming skills in Python. While students learn the technical theories and applications of AI, the use of responsible AI is emphasized throughout the course. The course was designed not only to help students assess and plan the human consequences of deploying the AI applications, but they also get to design and implement changes to mitigate or lower the risks associated in using these applications.

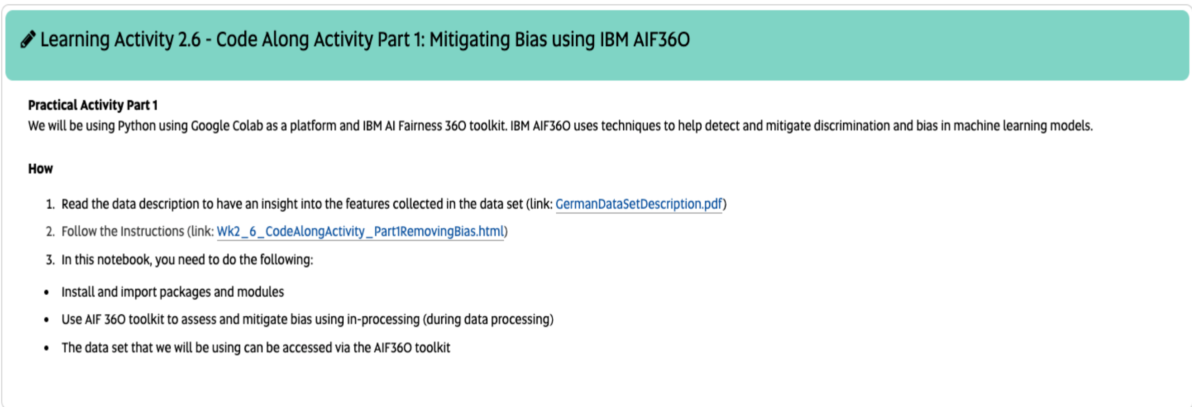
It was purposely decided that the principles of responsible AI will be the focus of the advanced data analytics courses where the technical practical ethics approach in teaching ethical issues in AI will be used. When students enroll in the course, they have a certain "maturity" in coding, they are no longer taught how to code. The expectation is that even with new code libraries introduced in the coding activities, students can learn and understand on their own. This is the reason why it was decided to embed the principles of responsible AI at the later part of the degree, so that students focus their attention on applying the principles of responsible AI in the topics that they are learning instead of learning how to code. Attention to the learned material is considered an important factor that influence learning [9].

3.1.1. Practical Ethics Applications

The course integrates responsible AI principles into data analytics education through a set of learning goals, formative activities, and assessments. Students are tasked with understanding and applying responsible AI principles, critically evaluating ethical issues in data and algorithms, designing mitigation strategies for AI-related risks, and effectively communicating ethical considerations. The curriculum employs a variety of formative activities, including code-along exercises, which include case study analyses and iterative problem-solving. All designed to reinforce responsible AI principles using a spaced repetition approach [10]. For instance, students analyze real-world scenarios like bias in credit scoring models [20], healthcare bias and data privacy issues [14], linkage attacks privacy issues [24], and data misrepresentation in public records [8]. They engage in critical code analysis, examining transparency, fairness, and privacy aspects of AI systems. Assessments are multi-faceted, comprising ethical impact analyses of real-world AI systems, practical implementation projects incorporating responsible AI principles, peer reviews of ethical reasoning, and reflective journaling. This approach, grounded in experiential learning theory [22], ensures students can bridge the gap between abstract ethical concepts and real-world applications in data analytics and AI development. By continually reflecting on the ethical implications of their work and iterating on solutions, students develop a robust understanding of how to apply responsible AI principles in practice, preparing them for the ethical challenges they'll face in their professional careers.

An example formative activity is the code-along activities (Figure 2). Students participate in guided coding exercises that incorporate responsible AI principles.

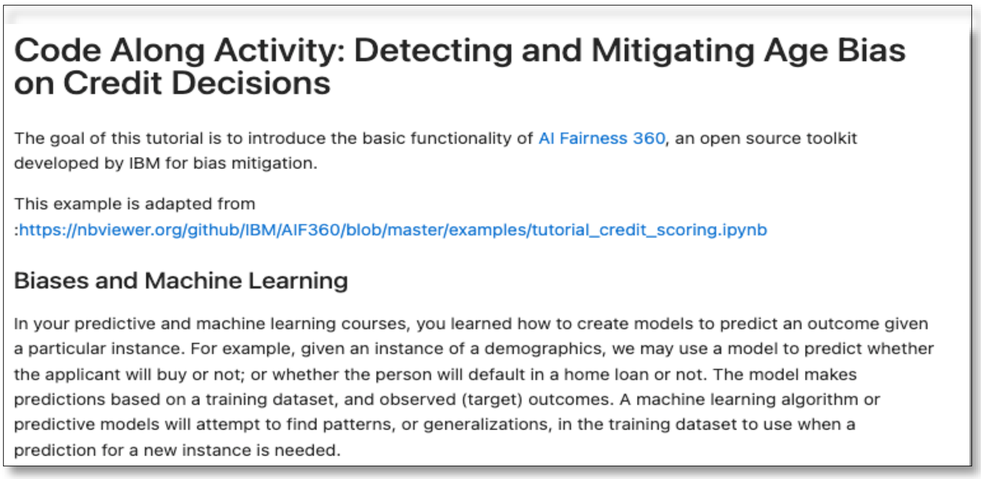
Figure 2. Example Code-Along Activity.



The Wk2_6_CodeAlongActivity_Part1RemovingBias.html in the activity contains the details of the walk-through of the code. In this example, students write a code and perform a critical observation during a code-along activity. Students learn how to use the toolkit to help enforce fairness and remove bias. Fairness metrics was used to check for bias in machine learning workflows, and bias mitigators was used to overcome bias in the workflow to produce a fairer outcome. Students are expected to apply their knowledge of responsible AI to assess these aspects critically.

As shown in Figure 3, students are introduced the tutorial (a), followed by sample code that they can code-along (b), and then some parts where students have to code (c).

Figure 3(a). Extract of code-along introduction.



Code Along Activity: Detecting and Mitigating Age Bias on Credit Decisions

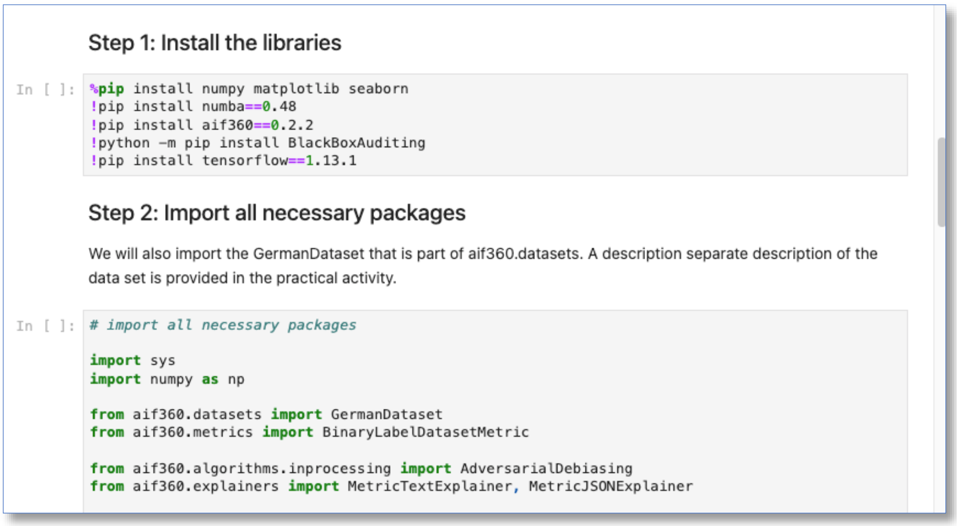
The goal of this tutorial is to introduce the basic functionality of AI Fairness 360, an open source toolkit developed by IBM for bias mitigation.

This example is adapted from :https://nbviewer.org/github/IBM/AIF360/blob/master/examples/tutorial_credit_scoring.ipynb

Biases and Machine Learning

In your predictive and machine learning courses, you learned how to create models to predict an outcome given a particular instance. For example, given an instance of a demographics, we may use a model to predict whether the applicant will buy or not; or whether the person will default in a home loan or not. The model makes predictions based on a training dataset, and observed (target) outcomes. A machine learning algorithm or predictive models will attempt to find patterns, or generalizations, in the training dataset to use when a prediction for a new instance is needed.

Figure 3(b). Extract of code-along instructions provided.



Step 1: Install the libraries

```
In [ ]: %pip install numpy matplotlib seaborn
        !pip install numba==0.48
        !pip install aif360==0.2.2
        !python -m pip install BlackBoxAuditing
        !pip install tensorflow==1.13.1
```

Step 2: Import all necessary packages

We will also import the GermanDataset that is part of aif360.datasets. A description separate description of the data set is provided in the practical activity.

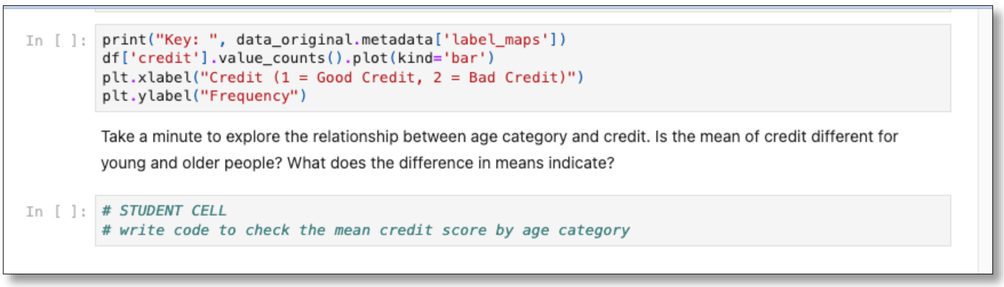
```
In [ ]: # import all necessary packages

        import sys
        import numpy as np

        from aif360.datasets import GermanDataset
        from aif360.metrics import BinaryLabelDatasetMetric

        from aif360.algorithms.inprocessing import AdversarialDebiasing
        from aif360.explainers import MetricTextExplainer, MetricJSONExplainer
```

Figure 3(c). Extract of code-along activity where students were asked to reflect and code.



```
In [ ]: print("Key: ", data_original.metadata['label_maps'])
        df['credit'].value_counts().plot(kind='bar')
        plt.xlabel("Credit (1 = Good Credit, 2 = Bad Credit)")
        plt.ylabel("Frequency")
```

Take a minute to explore the relationship between age category and credit. Is the mean of credit different for young and older people? What does the difference in means indicate?

```
In [ ]: # STUDENT CELL
        # write code to check the mean credit score by age category
```

After completing this activity, students reflect on and share this in the discussion forums:

- Ethical implications of bias in credit scoring
- Trade-offs between fairness and model performance
- Potential societal impacts of such models

The above example is the typical format of practical activities. Students are expected to repeatedly apply reflections on their comprehension of the data and the algorithmic fairness and diversity in data processing. Afterward, they reflect, plan, and utilize other principles of responsible AI that relate to the current topic. At the conclusion of each activity, students are requested to reflect on and discuss their feedback regarding some or all the principles, depending on their applicability. For instance, the reinforcement learning topic placed particular emphasis on transparency in addition to privacy and fairness. In contrast, discussion and reflections in the computer vision topic encompassed a broader range of principles, including reliability and safety.

The experiential learning concept was also adapted in the design of the assessment instruments. The criteria used for assessing students align with the course learning outcomes. Students are assessed in the different data analytics and AI concepts introduced in the courses. In each of these concepts, the assessments have the following format: a real-life case scenario is provided where students will be asked to write and submit a code and report. In the report, students are asked to explain the problem, discuss their approach to the problem, and discuss strategies in implementing the principles of responsible AI. The assessment was intentionally designed to require students to draw upon their understanding of abstract concepts in principles of responsible AI and apply that knowledge to their concrete experiences working with data, designing solutions, and engaging in actual coding. Additionally, students were tasked with resolving any conflicts they encountered between what they observed in the learning activities (e.g., code-along activities) and what they were doing (i.e., designing and coding solutions to problems). While fairness, diversity and privacy principles were integral components in all assessments) different principles were applied in various assessments. Figure 4 illustrates example instructions for assessing students' understanding and application of transparency and accountability in machine learning. In this assessment, students were presented with a case study scenario and tasked with designing and developing an optimized reinforcement learning model. Following this, they were required to discuss how their chosen reinforcement learning approach addresses principles of Responsible AI, specifically focusing on transparency and accountability. This exercise aimed to evaluate students' ability to not only implement advanced machine learning techniques but also to critically consider the ethical implications of their work within the framework of Responsible AI.

Figure 4. Example assessment instruction that focusses on the transparency and accountability principles.

5. **Transparency and Accountability**
    1. Discuss how the model addresses the transparency and accountability principles of responsible AI. For example, you may want to consider the following questions:
        1. Assuming that bias has been removed from the environment where the model learns, is the model designed to prevent bias to be introduced?
        2. Is the model transparent?
        3. Is it clear who will be accountable if bias is introduced?
    2. If the model is not addressing these issues, suggest steps that should be taken to address them in future developments of the recommender system.

4. Challenges

Integrating responsible AI principles into AI courses presents a multifaceted set of challenges that educators must navigate carefully. A primary concern is finding the right balance between technical content and ethical considerations. While it is crucial to prepare students with AI skills, there's an equally pressing need to instill a deep understanding of the ethical implications of these technologies. For instance, when teaching reinforcement learning, teachers must not only cover complex algorithms but also discuss potential biases in reward functions and their societal impacts. To address these issues, we integrated ethical discussions directly into technical topics using real-world examples, developed case studies combining technical implementation with ethical analysis, and implemented project-based learning in assessments.

Another significant challenge lies in maintaining the relevance of case studies and ethical scenarios in a rapidly evolving field. An example that was pertinent last year, such as facial recognition in public spaces, might be superseded by more pressing concerns like AI-generated deepfakes in political campaigns. We continue to search for current news articles to maintain relevance, though a more scalable approach is needed.

Assessing ethical reasoning poses its own set of difficulties. Unlike technical skills that can often be evaluated through quantitative metrics, judging the quality of ethical decision-making requires multifaceted criteria. Educators might struggle to develop rubrics that objectively measure a student's ability to identify and reason through ethical dilemmas in AI development. Assessing ethical reasoning posed unique difficulties, which we tackled by developing rubrics focused on the process of ethical decision-making.

Student resistance can also be a hurdle, as some may perceive ethical training as less valuable than technical prowess in the job market. This attitude might be reinforced by the tech industry's historical focus on innovation over ethical considerations, though this is gradually changing. To combat student resistance and demonstrate the importance of ethical skills, we highlighted job postings specifically mentioning ethical AI requirements.

Instructor expertise presents another challenge. Many AI professors come from technical backgrounds and may not feel equipped to lead discussions on complex ethical issues. Conversely, ethics professors might struggle with the technical intricacies of advanced AI systems. This gap necessitates either extensive cross-training or collaborative teaching models. The expertise gap among instructors was currently addressed by associating IT instructors with professional bodies offering ethics-related development opportunities. Finally, time constraints in already packed curricula can make it difficult to give both technical and ethical aspects their due attention. A course on natural language processing, for example, must cover a vast array of algorithms and techniques, leaving little room for in-depth discussions on the ethical implications of language models in areas like content moderation or automated customer service. While we've made significant progress, we recognize that this is an evolving process requiring ongoing adaptation. Addressing these challenges requires a thoughtful, multidisciplinary approach to curriculum design and a commitment to ongoing adaptation as the field of AI continues to advance and evolve.

5. Best Practices

Implementing best practices for integrating responsible AI principles into AI education also requires a thoughtful approach. Rather than treating ethics as a standalone topic, it's crucial to embed these principles throughout the entire curriculum. For instance, when teaching machine learning algorithms, teachers can consistently highlight potential biases in data sets and discuss the ethical implications of model choices. This integrated approach ensures that students view ethical considerations as an inherent part of AI development rather than an afterthought.

Utilizing real-world case studies is another vital strategy. For example, teachers might analyze the ethical concerns surrounding OpenAI's GPT models, discussing issues like potential misuse for disinformation, copyright infringement, and the amplification of biases. Such current and relevant examples make ethical dilemmas tangible and demonstrate their immediate relevance to the field. Hands-on ethical coding exercises further reinforce these principles. Students might be tasked with implementing fairness constraints in a credit scoring algorithm or designing transparency measures for a recommendation system, thereby gaining practical experience in translating ethical principles into code.

Collaborative learning plays a crucial role in developing a well-rounded understanding of AI ethics. Group projects could involve designing an AI system for a sensitive application, such as healthcare diagnostics, requiring students to collectively navigate technical challenges while addressing ethical concerns like patient privacy and algorithmic transparency. An interdisciplinary approach, possibly involving collaboration with ethics or philosophy areas can provide deeper insights into ethical frameworks and their application to AI. For instance, a joint seminar between computer science and philosophy students could explore the ethical implications of autonomous vehicles, bringing together technical knowledge and ethical reasoning. Continuous assessment of ethical reasoning is also essential to ensure that students are internalizing these principles. This could involve regular reflection papers on the ethical implications of topics covered in class, or project milestones that require ethical impact assessments.

6. Lessons Learned

The experience in embedding responsible AI principles into the AI technical course has resulted to valuable lessons that can significantly enhance the learning experience and outcomes for students. One crucial insight is the importance of timing in introducing these concepts. By incorporating ethical considerations into courses were students have the coding experience, students can focus on applying these principles to complex AI systems rather than grappling with basic coding challenges. For instance, in an intermediate level machine learning course, students might explore the ethical implications of using AI in criminal justice predictive systems, analyzing potential biases and fairness issues in real-world applications.

Practical application plays an important role in ethical AI education. When students can immediately apply ethical principles to their coding projects, engagement and understanding deepen significantly. For example, a project on developing a recommendation algorithm for a streaming service could include requirements for transparency in the algorithm's decision-making process and considerations for diverse representation in content suggestions. This hands-on approach allows students to see firsthand how ethical considerations shape technical decisions.

The development of ethical reasoning skills has been observed to require consistent practice and reflection. Regular opportunities for ethical decision-making, such as weekly case study discussions or ethical impact assessments for each major project, help students hone their ability to identify and navigate complex ethical dilemmas in AI. For instance, students might be asked to regularly update an "ethical journal or discussion forum" throughout the course, reflecting on how each new AI technique they learn could be used or misused from an ethical standpoint. In addition, relevance to current AI developments and potential career scenarios has proven to significantly enhance student engagement. Discussing recent controversies, such as the ethical implications of AI-generated art or the role of large language models in spreading misinformation, helps students see the immediate relevance of ethical considerations in their field. Another activity that has shown to be important in teaching responsible AI is peer learning. Group discussions on ethical dilemmas often lead to rich insights and perspectives that individual reflection might not yield.

Flexibility in curriculum design has emerged as a crucial factor, given the rapidly evolving nature of AI and its ethical challenges. Teachers must be prepared to adapt their teaching materials to address emerging issues, such as the ethical considerations surrounding new AI technologies. Assessment strategies have evolved to combine evaluation of technical skills with assessment of ethical reasoning. This might involve projects where the technical implementation is judged alongside an ethical impact report, providing a holistic view of a student's capabilities as a responsible AI practitioner.

Emphasizing the relevance of responsible AI principles to future careers has significantly increased student appreciation for these courses. When students understand how ethical considerations will impact their work in industry or research, they engage more deeply with the material. This could involve assignments that mimic real-world scenarios.

7. Conclusion and Recommendation

The rapid emergence and adoption of Artificial Intelligence technologies has led to increased opportunities, risks, and challenges, highlighting the need to teach the principles of responsible AI. In this paper, a case study was presented, showcasing the effective embedding of responsible AI principles in an advanced technical online course using the theory of experiential learning and applied ethics as a foundation. The course design incorporated practical and code-along activities that moved ethical discussions to specific topics, emphasizing the importance of understanding and checking data prior to modeling and evaluating models and algorithms, with the AI principles consistently integrated. The consistent presentation and reinforcement of the principles of responsible AI through the design of practical activities have proven beneficial in fostering students' comprehension and application of these concepts. It not only enhances awareness of the ethical challenges but also facilitates the practical applications of ethical concepts. Moreover, this approach fosters a conducive environment that stimulates critical thinking among students regarding ethical considerations and the far-reaching implications of their professional actions.

Despite the success of the course design, challenges were identified, such as finding relevant up-to-date case scenarios and lack of technical and responsible AI expertise among teachers. Developing practical activities and assessments for the course presented a significant challenge for the course designers. One of the main hurdles was finding case studies that were relevant and up to date with the latest technology-related events, followed by the task of modifying them to include the concepts being taught while also highlighting the principles of responsible AI. Additionally, finding suitable data to illustrate both the concepts and principles of responsible AI was also a challenge. In many cases, the data had to be modified to suit the purpose of the topic. Another challenge faced in the courses design was the lack of teachers who possessed both the necessary technical expertise and an understanding (or interest) in embedding responsible AI principles into their courses. This issue required extensive training and support to equip the educators with the necessary skills and knowledge to effectively integrate these principles into their teaching.

Best practices for addressing these challenges involve embedding ethics throughout the curriculum, utilizing real-world case studies, implementing hands-on ethical coding exercises, fostering collaborative learning, and adopting interdisciplinary approaches. Lessons learned highlight the importance of timing in introducing ethical concepts, the value of practical application, the need for consistent practice in ethical reasoning, the benefits of peer learning, and the necessity of flexible curriculum design. Additionally, emphasizing the relevance of responsible AI to future careers has been shown to increase student engagement and appreciation for these principles.

This research, while providing valuable insights into integrating responsible AI principles into technical courses, has several limitations and areas for future work. A key limitation is the lack of formal student feedback, which would provide crucial data on the effectiveness of the approach from the learners' perspective. Future research should prioritize collecting and analyzing student feedback to refine the curriculum design. Additionally, the study's scope was limited to a single course, potentially limiting its generalizability. Future work should focus on scaling up the integration of responsible AI principles across entire curricula, including non-AI related courses. This expansion would require developing comprehensive teacher training programs and establishing partnerships with industry to ensure ongoing relevance of case studies. Creating a dynamic, regularly updated database of ethical case studies and fostering multidisciplinary collaborations between computer science and philosophy departments could further enhance the approach.

Declaration on Generative AI

The author has not employed any generative AI tools.

References

[1] ABET, “Criteria for accrediting computing programs 2023-2024,” 2023, URL: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2023-2024/

[2] Association of Computing Machinery, “ACM Code of Ethics and Professional Conduct,” version 1,2023, URL: https://www.acm.org/code-of-ethics

[3] Australian Computer Society, “ACS Code of Ethics,” 2014, URL: https://www.acs.org.au/content/dam/acs/rules-and-regulations/Code-of-Ethics.pdf

[4] Google, “Artificial Intelligence at Google: Our Principles,” version 11, 2022, URL: https://ai.google/principles/

[5] D. Azim and M. S. Shamim, “Educational theories that inform the educational strategies for teaching ethics in undergraduate medical education,” Journal of the Pakistan Medical Association, 70.1(2020): 123-128.

[6] B. Blundell, “Ethics in Computing, Science, and Engineering: A Student's Guide to Doing Things Right,” 2020, Springer Nature, Switzerland.
[7] F. Bott, “Professional Issues in Information Technology,” 2nd ed. (2014), British Computing Society.
[8] City of New York, "DOHMH New York City Restaurant Inspection Results," NYC OpenData. URL: https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j.
[9] M. M. Chun and N. B. Turk-Browne, "Interactions between attention and memory," Current Opinion in Neurobiology, vol. 17 2 (2007):177-184, doi: 10.1016/j.conb.2007.03.005
[10] D. Ausubel and M. Youssef, "The Effect of Spaced Repetition on Meaningful Retention," The Journal of General Psychology, 73.1 (1965): 147-150, URL: https://doi.org/10.1080/00221309.1965.9711263.
[11] J. Davis and H. M. Walker, "Incorporating Social Issues of Computing in a Small, Liberal Arts College: A Case Study," in Proc. 42nd ACM Technical Symposium on Computer Science Education, 2011, pp. 69–74.
[12] M. Davis, Ethics and the University, Taylor and Francis, 2022.
[13] Meta AI, “Facebook’s five pillars of responsible AI,” version 11, 2022, URL: https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai.
[14] Fairlearn, “Fairlearn: A Python package to assess and improve fairness in AI,” version 11, 2022, URL: https://fairlearn.org.Ferrarello, L. (2019).
[15] L. Ferrarello, "Social awareness in design & engineering education and practice: the value of ethics in postgraduate education," in Proc. 21st International Conference on Engineering & Product Design Education, Glasgow, 2019, doi:10.35199/epde2019.83.
[16] C. Frauenberger, M. Rauhala, and G. Fitzpatrick, “In-Action Ethics, Interacting with Computers,” 29.2 (2016): 220-236, doi: 10.1093/iwc/iww024.
[17] K. S. Hansen, "Literature for Specific Purposes: A Literary Approach to Teaching Ethics in Science and Technology," Configurations (Baltimore, Md.), 26.3 (2018):337-343, doi: 10.1353/con.2018.0030.
[18] A. Harris, M. Lang, D. Yates, and S. E. Kruck, "Incorporating Ethics and Social Responsibility in IS education," Journal of Information Systems Education, 2.3 (2011):183-190.
[19] J. Herrington, R. Oliver, and T. Reeves, "Patterns of engagement in authentic online learning environments," Australian Journal of Educational Technology, 19.1 (2003): 59-71, doi: 10.14742/ajet.1701.
[20] H. Hofmann, "German Credit Risk," 2016, version 2, 2022, URL: https://www.kaggle.com/datasets/uciml/german-credit.
[21] H. Kathuria and D. W. Becker, "Leveraging Course Quality Checklist to Improve Online Courses," Journal of Teaching and Learning with Technology, 10.1 (2021), doi: 10.14434/jotlt.v10i1.31253.
[22] D. A. Kolb, "Part 1: Experience and Learning,” Experiential Learning: Experience as the Source of Learning and Development, 2nd ed., Pearson Education, 2014.
[23] S. Kolber, "Enhancing accessibility with instructional video," Teaching Online Master Class, 2020.
[24] P. Ohm, "Your secrets live online in databases of ruin," Ars Technica, (2009). URL: https://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/.
[25] J. Saltz, M Skirpan, C Fiesler, M Gorelick, T Yeh, R Heckman, N Dewar, Integrating Ethics within Machine Learning Courses, ACM Transaction on Computing Education, 19.4 (2019) 1-26. doi:10.1145/1188913.1188915.
[26] Microsoft, Microsoft Responsible AI Standard, v2, version 11, 2022. URL: https://www.microsoft.com/en-us/ai/responsible-ai
[27] IBM, AI Ethics, version 11, 2022. URL: https://www.ibm.com/impact/ai-ethics
[28] Accenture, Responsible AI : AI Ethics & Governance, version 11, 2022, URL: https://www.accenture.com/au-en/services/applied-intelligence/ai-ethics-governance
[29] Open AI, OpenAI Charter, version 10,2022, URL: https://openai.com/charter/

[30] D-W. Olayinka, Teaching Ethics in Decision Making: Embedding Moral Reasoning in the Management of Information Systems, Teaching Ethics Across the Management Curriculum, K. Ogunyemi, Ed. New York: Business Expert Press, 2014, pp. 179-197.
[31] A. Scott and S. Barlowe, "How software works: Computational thinking and ethics before CS1," in Proc. IEEE Frontiers in Education Conference (FIE), 2016, pp. 1-9, doi: 10.1109/FIE.2016.7757436.
[32] M. Skirpan et al., "Ethics Education in Context: A Case Study of Novel Ethics Activities for the CS Classroom," in Proc. of the 49th ACM Technical Symposium on Computer Science Education, 2018, pp. 940-945.
[33] I. Stavrakakis, D. Gordon, B. Tierney, A. Becevel., E., Murphy, G. Didg-Crnkociv, D. O'Sullivan, "The teaching of computer ethics on computer science and related degree programmes: A European Survey," Int. J. Ethics Educ., 7.1 (2022): 101–129, doi: 10.1007/s40889-021-00135-1.
[34] T. Stratton, "Incorporating ethics content throughout an integrated pharmacy curriculum," Am. J. Pharm. Educ., 84.11, 2020, doi: 10.5688/ajpe7865.
[35] SUNY Online, "The SUNY Online Course Quality Review Rubric (OSCQR)," version 09, 2022, URL: https://oscqr.suny.edu/.
[36] C. Wankel, "Design considerations," in Redesigning Courses for Online Delivery Design, Interaction, Media, and Evaluation, vol. 8, R. Parker, Ed. Emerald Group Publishing Limited, 2013, doi: 10.1108/S2044-9968(2013)0000008003.
[37] D. E. Wueste and D. Scibilia, "Cognition and Conation: A Potent Alliance in Teaching Ethics," in Instructional Models, Methods, and Modalities for University Studies, pp. 5-20, Rowman & Littlefield Publishers, 2021.
[38] R. Zhang, N. C. Bi, and T. Mercado, "Do zoom meetings really help? A comparative analysis of synchronous and asynchronous online learning during Covid-19 pandemic," Journal of Computer Assisted Learning, 9.1 (2023): 210–217, 2023, doi: 10.1111/jcal.12740.
[39] N. Lee and S. J. Kim, "A Systematic Course Design Approach to Guide the Development of a Construction Engineering and Management Capstone Course," in 2020 ASEE-NE Annual Conference, pp. 1–5, doi: 10.1109/ASEENE51624.2020.9292652.
[40] A. Lewis and J. Stoyanovich, "Teaching Responsible Data Science: Charting New Pedagogical Territory," Int. J. Artificial Intelligence in Education, doi: 10.1007/s40593-021-00241-7, 2022.
[41] J. Macdonald and A. Campbell, "Demonstrating online teaching in the disciplines: A systematic approach to activity design for online synchronous tuition," Br. J. Educ. Technol., 43.6, (2012): 83–891, doi: 10.1111/j.1467-8535.2011.01238. x.
[42] F. Martin, A. Ritzhaupt, S. Kumar, and K. Budhrani, "Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation," Internet High. Educ., 42, (2019): 34-43, doi: 10.1016/j.iheduc.2019.04.003.
[43] S. J. McGhan, C. M. Jackson, and K. Premer, "Online course quality assurance: Development of a quality checklist," InSight: A J. Sch. Teach., 10.1 (2015): 126-140.
[44] O. Poquet, L. Lim, N. Mirriahi, and S. Dawson, "Video and learning: A systematic review (2007–2017)," in Proc. of the 8th Int. Conf. on Learning Analytics and Knowledge—LAK, 2018, doi: 10.1145/3170358.3170376.
[45] "Ethics of Artificial Intelligence," UNESCO.org, 2023. URL: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.
[46] "Ethics guidelines for Trustworthy AI," Shaping Europe's digital future, 2019. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.