Applied Ethics via Encouraging Intuitive Reflection and Deliberate Discourse

Lucas J. Wiese, Purdue University

Lucas Wiese is a PhD student in Computer and Information Technology at Purdue University. He studies AI ethics education and workforce development and works in the Research on Computing in Engineering and Technology Education lab (ROCkETEd) and the Governance and Responsible AI Lab (GRAIL).

Dr. Alejandra J. Magana, Purdue University

Alejandra J. Magana, Ph.D., is the W.C. Furnas Professor in Enterprise Excellence in the Department of Computer and Information Technology and Professor of Engineering Education at Purdue University.

Applied Ethics via Encouraging Intuitive Reflection and Deliberate Discourse

Abstract

Artificial intelligence's (AI) widespread societal impact means that students of all disciplines will be working in roles adjacent to this new technology. As a result, they need to understand how to appropriately navigate and behave ethically in practice. The purpose of this paper is to introduce and detail a learning intervention intended to enhance the ethical behavior of future AI developers and engineers. The SIMDE conceptual framework was developed to offer a basis for understanding the pre-rational aspects of ethical decision-making as they are carried out into deliberate discourse in a social space amongst peers. To investigate the SIMDE framework, students were asked to solve a professional AI ethics problem in a dilemma-based seven-step learning activity. The qualitative results of this paper examine how constructs in the SIMDE conceptual framework were present in student responses, and what students learned from peer discourse that led them to either justify their gut-reaction decision or change their mind. We found that students are impacted by perspective-taking, they use reasoning to defend their position rather than seek and appraise truth, and moral self-reflection helps them learn more about themselves. Moreover, even when students learn new information and improve their reasoning, they are not inclined to change their minds from their initial intuitive judgment. This finding supports literature that suggests 'reasoning' can only go so far in the ethics curriculum if behavioral change is the goal. More interdisciplinary educational research is necessary to design an ethics curriculum that can appropriately prepare future AI professionals for the demands of industry.

1. Introduction

This evidence-based practice paper details a novel learning intervention for applied ethics education curriculum that leverages students' intuitions as a precursor to the ethical decision-making process. In 2004, Bertolami voiced a concern that ethics is boring: "Most ethical principles are simply too abstract, dry, and off-putting to have any practical effect." [1, p. 417]. Yet, as artificial intelligence (AI) becomes more pervasive, and more students outside of engineering and computer science disciplines will work with AI, ethics education becomes increasingly critical.

Fortunately, scholars have worked to update ethics curricula in engineering and technology departments to attempt and make ethics more practical. And, in AI curriculum specifically, ethics are often "embedded" in the nature of the AI development lifecycle from data considerations and sociotechnical effects of unchecked bias in AI models [2], [3], [4]. Unfortunately, ethics still seems to be cast away as unimportant and disconnected from the reality of work [5]; Moreover, the practical ethical skills that students learn may not lead them to the societal outcomes we want [6].

There is some debate about whether ethics is the responsibility of a technical curriculum or an interdisciplinary effort from the university [2], [5], [7], [8]. While it is likely that benefits can come from multiple parties taking responsibility for ethics education, the goal of ethics education is also under debate [6]. Many progressive ethics curricula, beyond those that just review theories and codes of ethics, have focused on *reasoning* as an updated and noble pursuit [6], [9], [10], [11]. However, even if ethical reasoning practice is grounded in real-life scenarios, recent evidence suggests that *action* is more linked to emotion, identity, moral engagement, and intuition [12], [13], [14]. Higher ethical *reasoning* skills may not lead to more positive ethical action [15] and thus calls for reformatted ethics education goals [11]. As a result, the ethics curriculum faces a challenge to foster an environment where students learn how to make decisions with positive ethical outcomes when working in technical environments. This paper details the learning design of an educational intervention intended to enhance the ethical behavior of future AI developers and engineers. By exploring how personal aspects of morality impact professional ethics scenarios, students may better understand ethical requirements in organizations and behave more ethically as a result.

2. Background

Education about AI, for both technical and nontechnical students, often includes an ethics component [16], [17]. However, the content of this ethics component can vary from focuses on technical constructs like privacy, transparency, and explicability, to focuses on social constructs like inequality, justice, and human autonomy [3], [18]. And while this ethics content is often taught in a traditional rote-knowledge-centric manner, rather than applied practice-based methods [19], [20], there have long been attempts to study and improve ethics curriculum in a practice-oriented way [21], [22]. However, despite good intentions, these applied ethics education strategies may still carry misguided learning goals that do not reflect foundations from moral psychology [11]. Specifically, a misguided aim that *reasoning* is touted as a first-mover toward enhanced ethical behavior and outcome from students.

While an extensive review of the evolution of ethics education is outside the scope of this paper (see [22]), there have been a few noteworthy strategies to implement applied ethics in technology and engineering-adjacent classrooms. For instance, Bairaktarova and Woodcock [23] developed a motivational model to emphasize a path toward behavior, with implications for pedagogical implementation. Hess et al. [10] conducted a five-semester study to examine pedagogical strategies for increasing ethical reasoning and empathic perspective-taking. And Fox and Beiter [24] leverage intentional failure to reach more critically appraised ethical decisions in engineering. Building from these strategies, there have then been calls to move toward ethical behavior and ethical action as the necessary goals for ethics education [6], [25], [26].

While achieving ethical action is complex, the following steps give a high-level descriptive account of how ethical action is reached. First, an eliciting situation triggers moral intuitions (including emotion). Then, an a priori intuitive judgment is formed. Following that, post hoc constructions work to reason and justify the intuitive judgment. Then, from a combination of personal moral beliefs and values, ethical action is taken in accordance with the judgment [13], [27], [28].

3. SIMDE Conceptual Framework

This study was motivated and guided by the Social Intuitionist Model (SIM) [13] and Jürgen Habermas's theory of Discourse Ethics (DE) [29], [30]. While more detail on each theory is expanded on in subsequent sections of this paper, and the full philosophical defense of each theory is out of scope, the key contributions of each theory follow. The SIM provides us that intuitions are the primary link (rather than reason) in the resulting judgment, and subsequent decision, from an eliciting situation. DE provides us grounds that say the only valid ethical norms come from agreed and consensual peer discourse. Refer to Fig. 1 to see how these two theories are connected in the classroom learning intervention.

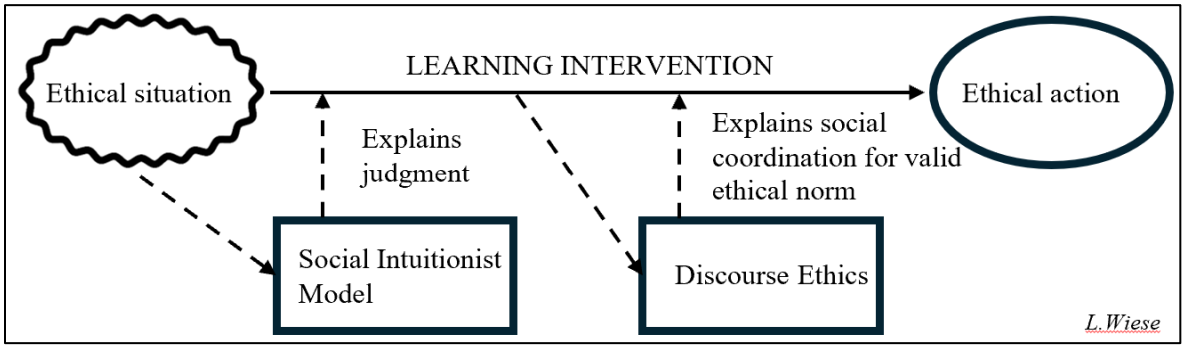

Fig. 1. Connection of SIM based learning intervention design to elicit ethical action from ethical dilemmas with DE.

The foundations of the SIM and DE originate in two different fields (moral psychology and social theory) but both maintain compatible philosophical origins. Prima facie, they seem contradictory: DE says that moral principles are derived from rationality, whereas SIM says that moral judgment comes primarily from non-rational cognition. However, both maintain cognitive origins of morality; where morality *can* express propositions that are true or false. It is a matter of sequence; SIM understands the cognitive formation of morality (pre-rationality to judgment) and DE understands the post-judgment rational procedures to form moral principles. For the remainder of this paper, I will refer to this as the SIMDE conceptual framework. And while not a necessary component of this conceptual framework, the political theory of Deliberative Democracy (DD), also originating from Habermas, helps reify the resulting moral principles from DE and brings an emphasis to the practical implications of decision-making [31].

Together, the SIM first provides an underlying descriptive account of how ethical decisions are formed, and second, DE informs pedagogical implications for talking about ethics in group discourse; First accepting psychological limitations of ethical decisions, and second, leveraging rational discourse to overcome individual limitations to form valid ethical norms. The following section details the learning intervention and how this SIMDE conceptual framework was pedagogically implemented.

4. Design for the learning intervention

The classroom environment is organized around two principles. First, inspiration from the Social Intuitionist Model for moral judgment de-emphasizes private reasoning and instead highlights that intuitive feelings and moral attitudes toward situations are the predominant factors that impact decisions. Second, as social and cultural influences sway decision-making more than private reflection, the classroom is organized around Discourse Ethics. Through deliberation and communicative action, this theory provides a procedure for deriving concrete ethical norms from abstract concepts. In this learning design, students internally investigated their moral tensions about an ethical dilemma and practiced deliberative discourse among peers to understand the complexity and multiplicity of ethical decision-making.

Before the learning intervention began, students were given a pre-reading assignment as homework a week before the classroom activity. In this homework, students reviewed the Markkula Center for Applied Ethics' framework for ethical decision-making [32] which gave a basic overview of how ethical theories can impact decision-making, and students were given a list of ethical principles that apply to AI in society [33], [34]. Short introductory video lessons about AI were also given to students, provided by UNESCO [35]. These short assignments before the learning intervention helped ensure students had similar levels of base knowledge about AI, ethics, and general decision-making processes. The following subsections describe the learning intervention chronologically as the activity progressed in class. These sub-sections describe the activity and then pedagogical elements about each stage's pedagogical benefit and connection to the SIMDE conceptual framework.

4.1. Dilemma

The learning activity began with a dilemma, which all other questions and decisions were derived from. The students were read the dilemma in class and given a digital handout at the same time. It was important that they did not see the dilemma before the learning activity began. The dilemma (named *Alaina's Dilemma*) was concerned with a hypothetical AI developer, Alaina, at an AI-based real-estate firm. In the fictional case, she developed a product that would identify ideal locations for real estate investment, reduce construction costs, and time and resource requirements. Moreover, she was under pressure from her financially struggling company to release a beneficial product. The problem was that she found, at the last minute, that the product contains substantive biases that disadvantage low-income areas and favor affluent high-income buyers—leading to further inequality in the (un)affordable housing crisis. She had to choose between releasing the product immediately for short-term financial gains (job stability and company growth) or investing more in the product to balance biases and fairness metrics in the algorithm (societal well-being). Students were asked whether they would choose the former, **option A**, or the latter, **option B**. With written instruction and verbal reminders, the students were asked to write down their gut reaction—their instinctual choice—without intentionally reflecting or reasoning.

Teaching ethics purely by theory or principles cannot appropriately account for the complexity that professionals face in practice. [20], [36], [37], [38]. Thus, case-based pedagogy attempts to reduce the abstraction found in theory yet tame the complexity found in practice to be useful in the classroom. The use of cases in education has long been used and can center learning around

specific real-world scenarios; And is in line with the "first" principle of instruction, "learning is promoted when learners are shown the task that they will be able to do or the problem they will be able to solve as a result of completing a module or course." [39, p. 45]. Moreover, the use of case-based pedagogy in *ethics* education has been linked to increased decision-making abilities and sensitivity to recognizing ethical issues [40], [41]. Similarly, grounding the learning in a *problem* rather than traditional lecture formats promotes moral and ethical development in students [42]. Now, the use of a dilemma is a key tenet in this study.

The intention of the 'dilemma' as a first step is to create moral tension and ambiguity surrounding a "correct" answer. This can simulate that situations are not black-and-white, and can promote more critical analysis and judgment about the case rather than defaulting to an "easy" answer [36]. This tension in the dilemma provides a root to investigate and unveil the conflicting moral values in everyday decisions. In provoking these moral tensions, the emotions surrounding a case can be more apparent—activating an overlap between emotional cognition and rational cognition—which can both help students learn and also help transfer skills from practice to real life [43]. Moreover, as Johnson et al. [41] found, cases comprised of more mixed and negative outcomes may improve students' ethical decision-making and ability to make ethical decisions. While the dilemma in this present study was not weighed toward any positive or negative outcomes, the lack of clarity for a positive outcome may encourage students to think more critically about the case. Centrally, hinging the learning activity on a dilemma-based case provides a stage to deepen critical thinking, elicit moral intuitions, and transfer ethics in practice to real life.

4.2. Intuition inventory

Following the dilemma, the students were asked to turn to the in-class digital worksheet that contained all the relevant questions and prompts for this learning intervention. First, students were asked to write down the emotions, feelings, or intuitions that they believed led them to their decision. It was encouraged that students try not to write down rationales or reasons for their 'A' or 'B' decisions. Moreover, in this quasi-a priori stage of the activity, students were asked to pick, from a list, which ethical principle they felt most at stake.

The recognition of intuition and its role in how judgments are formed, decisions made, and actions taken is the central component of this learning activity. Primarily leveraging the *social intuitionist model* (SIM), intuitions are the primary link between an eliciting situation and the formation of a judgment (an evaluation of good or bad). Intuitions hold an invisible weight over reasoning and sway reasoning [13], [28]. Teaching students—and professionals—to recognize their intuitions can help them more critically evaluate the weight of their actions in a dilemma [11], [28], [44]. And to Merrill's [39] instructional design principles, recognizing intuitions may help "activate" learning by connecting the problems in the dilemma to their past experiences as individuals. Moreover, it might promote "integration," or learning transfer, as students will associate future cases' emotional valence with the practice case studies' emotional and intuitive valence [39], [43]. While beyond the scope of the paper, more literature related to morality and ethics education about how individuals (e.g. students) are swayed in ethical decision-making can be found in the theory of bounded rationality [45] and nudge theory [46].

4.3. Establishing knowledge and sociotechnical fact-gathering

After students were prompted to consider a priori elements of the decision-making process, there was a brief interlude in the activity where the instructor briefly introduced the ethical decision-making process and how intuitions and emotions sway common actions and decisions. Now, students were asked to consider the "facts" of the case. Here, students simply enumerate any relevant information they thought was relevant to the decision between option 'A' or 'B'. The worksheet prompted them with guiding questions, too, such as: *Who are the people affected? What groups are most affected? What values, principles, laws, or regulations might be most at breach? Or, most important to uphold? How does AI technology impact these decisions?* At the end of this stage in the activity, students were asked to define the ethical principle they chose in their own words, without computer aid, and with an anecdote from their own life.

This stage begins pedagogical alignment with Kisselburgh et al.'s [47], and Hess et al.'s [10] extension, scaffolded, integrated/interactive, and reflective analysis (SIRA) framework. This stage expects students to consider the "basic facts" of the case—establishing a grounding of knowledge about the sociotechnical space surrounding the dilemma. In this stage, students are prompted to specify (give a rudimentary definition for) an ethical principle that they *identified* in the previous stage. This formulated a more formal procedure for ethical reasoning, based on Beever and Brightman's [48] Reflexive Principlism approach. Moreover, this procedure of operationalizing an ethical principle as students gather sociotechnical knowledge about a case can set the stage for rational discourse [49].

4.4. Small group discussion

Once students privately reflected on their intuitions and gathered facts of the case, they were now encouraged to speak amongst small groups. By the time of the learning intervention, students were already used to working with certain peers in a small group fashion, so students paired with these same individuals (i.e., students selected their own groups). In these groups, students discussed whether they chose 'A' or 'B', and what reasons or instincts led them to their choice. Moreover, students were asked to go over what facts they thought were relevant and what ethical principles were under consideration. Near the end of this stage, students were asked to jot down a few things they found interesting after group discussion and whether they changed their mind on the option 'A' or 'B'.

As students entered discussion with their peers, a common practice in the classroom, there were two main intentions. First, they were introduced to alternate perspectives and could begin perspective-taking to compare and contrast different viewpoints and arguments surrounding the dilemma [10]. Moreover, "the most widely discussed method of triggering new intuitions is role-taking… Simply by putting oneself into the shoes of another person, one may instantly feel pain, sympathy, or other vicarious emotional responses." [13, p. 819]. Second, a central tenet of this learning intervention is to rely on—and emphasize—group discussion and discourse in the decision-making process. Based on the limitations of the individual (not being as rational and objective as we would like), the theory of Discourse Ethics supposes a procedure to procure valid ethical norms based on the inclusion and voice of every participant relevant to a problem [50]. While this in-class activity cannot achieve such reach, the practice of including others in discourse when faced with morally ambiguous dilemmas can foster more responsible decision-

making. Discourse Ethics has been applied to organizational communication and business ethics [29], [30], [49] as well as in-classroom pedagogy [51], [52] to achieve more responsible and improved ethical development.

4.5. Expert opinion and hands-on demonstration

To push students outside of personal beliefs and peer opinion, this stage introduced students to (a) expert opinions and (b) real machine learning code relevant to Alaina's Dilemma. First, the expert opinions came from accessible public-facing news outlets that reported on the manner: MIT Technology Review [53] and The Lever [54]. Second, the machine learning code was demonstrated to the students during class. This was an end-to-end machine learning project [55], from downloading and cleaning the data to training and refining the model, which was modified for students' skill level by the instructor.

Introducing outside perspectives and expert opinions helps induce "safe" conflict in the students' beliefs and perspectives. Here, they further learn to consider outside information in their decision-making process and that ethical dilemmas are more than black-and-white problems [10], [47]. This may increase "moral tension" in the students (as discussed in the introduction of the 'Dilemma') and further increase ethical awareness and sensitivity. Moreover, in this stage, real ML code was demonstrated to students so they could see how these problems apply in practice [39]. Demonstrating real practices to students can help "make it real" for students and show how practice in the classroom can translate to real life.

4.6. Classroom wide discussion

Students then had the chance to discuss classroom-wide (n = 96). Instead of just between their peers, the instructor facilitated the discussion so that students could share their stance about choice 'A' or 'B', and why, to encourage moderated discourse and conflict. Students were encouraged to disagree and rebuttal. This resulted in visual "ah-ha" moments for some students as they heard counter-arguments to their beliefs. This stage, paired with the previous stage, fostered conflict and discourse to go beyond individual and group influence.

In addition to the pedagogical benefits of discussion and conflict, mentioned prior, this stage introduced the possibility for "failure" more than previous stages. For instance, students (a) had to hold a level of self-confidence to speak about their decision to the class and (b) accept that it might *not* be accepted by the class, or that others will disagree. Specifically, this stage can draw from work on failure-based learning, where "failure" is encouraged in the classroom and fosters growth. By accepting some "failure," students can begin to understand that complex problems take time and iteration to fully understand and that there will be conflict and differences in a problem-solving environment [24], [56]. In this, the hope is that students accept risk marginally more than before; Here, students can recognize it is okay to speak up about something they believe in and take a risk that others may not be as receptive. In turn, this stage allows students to give a voice to their values. In line with the Giving Voice to Values (GVV) [57] pedagogical approach, students practiced post-decision-making skills and further developed students' willingness to express their stances and positions on a decision.

4.7. Final small group discussion and conclusion

In the final stage of the in-class activity, students returned to their small groups and were asked to "balance" their ethical principles in one paragraph. While individual students may have selected different ethical principles as most relevant, they could talk to others if necessary to complete the identification, specification, and balance of an ethical principle in a certain context. Next, students were prompted to write an argument to defend their final decision (option 'A' or 'B'). In the activity instructions, it was noted that this "argument" could be similar to what may need to be said to a manager or board of a company about why option 'A' should be pursued versus 'B', or vice versa.

As mentioned, students completed the formalized procedure for ethical reasoning as students balanced the ethical principle among real-world constraints and alternate perspectives from peer discourse [10], [48]. This both implicitly showed that ethical principles need to be contextualized and provided a site for data collection for future studies on ethical reasoning in a Reflexive Principlism-integrated pedagogical framework. As students wrote an argument, this further built communication skills, self-efficacy, and confidence. While it is a simple post-decision-making step, the cultivation of strong communication skills is often missed in courses that promote ethical development, ethical reasoning, or basic ethics knowledge [25].

4.8. Private take-home reflections

To complete the learning intervention, students were asked to complete a take-home reflection assignment. Students reflected on: (i) what they believed impacted a change in their thinking the most, (ii) if discussing with others helped them justify their position more or change their mind, and (iii) what about AI or advanced technologies uniquely impacts these decisions and thinking processes. In addition to these three reflection prompts, general feedback was gathered about whether the activity could have been clearer if they (honestly) learned anything and a question for "anything else." Students submitted this take-home reflection assignment, and the learning intervention was completed.

Reflection in learning has been widely recognized as an essential element of learning. As students reflect, they consciously explore what they have learned and discover new aspects of their thinking [58]. Moreover, reflecting on metacognitive processes is indicative of "better learning" whereas the lack of reflection can lead to unsuccessful learning [59]. While the use of reflection is essential to learning and internalizing the learning activity, reflection is also essential to triggering new intuitions—and recalibrating old intuitions [13], [28].

5. Methods

This learning intervention took place in a junior-level course in the Computer and Information Technology (CIT) department titled "Policy, Regulation, and Globalization in Information Technology" during the Fall 2023 semester at a large midwestern university. The deployment of the intervention took place during one lecture in the latter half of the semester. While 96 students were enrolled in the course, not all students attended the lecture on that day, did not submit the necessary reflection assignment, or submitted near-blank responses. Therefore, the total sample

size of this study consisted of 70 undergraduate CIT students. Demographic data such as age, gender, or year in school were not collected and thus not reported.

During the full learning intervention, students submitted three assignments (a pre-, during-, and post-activity assignment) along with a quantitative pre-post survey to measure intent to behave. The dataset for this present study was limited to the third assignment, the post-activity reflection assignment (n = 70 observations). There were six questions in this assignment, and while considered holistically at times, only three of the questions were in focus for this study. First, data were extracted from student submissions and prepared in an Excel spreadsheet. Here, a document-level (all questions per student) code was applied to assess student learning on a basic qualitative self-report question (see Results). Then, data were imported to ATLAS.ti 23.4.0 [60] to organize the data, facilitate coding, and assist with analytic insights from the dataset. To begin analyzing the data, the research team followed a basic thematic analysis [61], [62] and began by reading the student responses while taking notes on potential codes and code categories. Then, a codebook was created that represented student responses in respect to the principles of the learning design. Each student submission was read and codes were applied. See the next section for reporting on the data analysis and more detail about the questions that were coded and the codes themselves.

6. Results

6.1. Feedback on intervention and classroom feasibility

First, this section establishes a basic understanding of how students responded to the intervention. In the take-home reflection homework assignment, students responded to the question: "Honestly, do you think this class taught you anything?" Student responses were coded in the three following categories, with frequency in parentheses: 'yes' (58), 'no' (7), and 'maybe' (5).

In these open-ended answers, students often expanded on the question and gave detail for why they answered 'yes', 'no', or 'maybe'. For instance, the following two quotes are from students who answered 'yes': "*Yes, it was interesting to compare my intuitions with my reasoning*," and, "*This class has really taught me to be more mindful of ethical decisions I make in my career endeavors*." Moreover, the following two quotes showcase students who answered 'no': "*No, but that's because I have taken other courses on ethics*," and, "*Unfortunately, I didn't really learn anything mind boggling or different.*" Last, students who were coded for 'maybe' either answered ambiguously or did not answer the question. For instance, "*I did gain knowledge that it is important to talk to others when making an ethical decision, [but] I personally do not know what else I have learned from the lecture*," and, "*I would have preferred a standard lecture about AI [ethics]."*

6.2. Evidence of SIMDE in the learning intervention

This section examines how constructs in the SIMDE conceptual framework were present in student responses. Students responded to the following question in their reflection assignment: "What do you think most impacted any changes in your thinking, or helped you clarify what you initially believed?" The open-ended answers to this question were coded for explicit and implicit

references to the elements in this learning design; the following constructs, with frequency counts in parentheses, were found: ‘peer discussions’ (26), ‘class discussions’ (10), ‘reflection’ (19; 5 basic reflection, 14 moral reflection), ‘fact gathering’ (9), ‘other specific element of learning activity’ (20). These codes were not mutually exclusive. For instance, some students mentioned that ‘peer discussions’ and ‘class discussions’ were the most impactful, or any combination of the codes, and thus those responses were double-coded. Refer to Table 1 for representative student responses for each code category.

TABLE I Students’ Reflections on the Conceptual Framework*

Code category	**Representative quote**
Peer discussions (26)	• I think the greatest change to my beliefs developed after discussing with my group members and other group mates. Hearing differing perspectives helped to grow my frame of view and expose me to different moral perspectives. • I think communicating thoughts and perspectives with others helps with the thinking of possible changes, but also stabilizing what I believed. • *Having time for peer discourse and a primed environment for ethical discussions helped students critically think about ethical decisions that they otherwise would not have—whether that led them to more confidence in their values or more openness to other perspectives.*
Class discussions (10)	• Feedback from the class definitely impacted any changes in my thinking along with the group discussion. • I think what impacted my thinking most [was] when the class discussed the consequences that might befall [other] stakeholders in the scenario. • *Taking discussions out of small peer groups further confronted individual beliefs and ethical decisions were being formed from a more expansive pool of perspectives.*
Basic reflection (5)	• I think the way I perceive reality, how I was raised, and who I am as a person and my surroundings. • *Reflective thinking, even when not explicitly about morality or values, helps students see themselves in relation to others and can promote empathic perspective-taking in ethical reasoning*
Moral reflection (14)	• My thinking changed when I reflected on my personal beliefs and ethical values. • *The emphasis on recognizing intuitions during the activity helped students reflect on their beliefs and value systems and helped*

	change—or inform—paths of thinking in the ethical decision-making process.
Fact gathering (9)	• I believe making small personal research on the specified topic helped me clarify my initial beliefs. • *Elements from the learning design scaffolded the fact-gathering process. This helped provide structure and evidence for a priori beliefs and knowledge about the case.*
Other element of activity (20)	• I liked the transition between individual reflection to small group discussion to the whole class. • I think what most impacted my thinking was the ethical dilemma that was presented to the class. • Being placed in an actual scenario where ethical decision-making [was necessary] helped put things into perspective and made me challenge my thinking. • *The intentional structure of the learning design helped students absorb this non-traditional applied ethics module and the "uncomfortable" elements of making difficult decisions helped push students to see outside of a black-and-white box.*

* Student quotes are in place text. Each code category contains an *italicized bullet point* with an interpretive analysis of the salient meaning of the quotes.

6.3. Factors in peer discourse that change or justify thinking

To further understand the role of discussion (peer discourse) on students' ethical decision-making and judgment, students responded to the following question: "Did discussing with other people help you justify your position more, or did discussing with other people help you change your mind at all?" These responses were each coded for (1) whether discussions changed the student's mind and (2) what *about* the discussion was mentioned. As a result, 53 students indicated "no, discussions did not change my mind," and 14 students indicated "yes, my mind was changed after discussions with others." Then, students' responses included four main categories for what is was "about" discussions that were most salient. These were, "discussions helped me **reason** and argue my position better," "discussions gave me new insight into **perspectives**," "discussions gave me new **knowledge**," and "discussions helped, in **general**."

While students who mentioned 'knowledge' and 'general' discussion elements were split evenly (with respect to the ratio of students who changed their mind, 'yes', to those that did not, 'no'), the code categories for 'perspectives' and 'reasoning' were not distributed in proportion to the total responses. For instance, while 21% of the students indicated 'yes' (discussions helped change their minds), 54% of the students who focused on 'perspectives' indicated that discussions changed their minds. Moreover, not a single student who focused on discussions helping them reason indicated discussions changed their mind; I.e., all students who got

reasoning skills out of the discussion did not change their minds, but rather, solidified and justified their existing position.

For the students who indicated they changed their mind, they most often referenced the effect of 'perspectives' in that they could see new perspectives and hear others' points of view. For instance, one student said, "*discussing with other people helped me to change my mind, as I could hear the opinions of an entirely different point of view. The discussion helped me see things that I would have overseen when thinking about the dilemma by myself*." The second most mentioned focus was that students gained new knowledge or information about the case through discussions. For instance, one student said, "*Discussing with other people helped change my mind. It opened me up to new possibilities and implications of unethical AI usage*."

However, as mentioned, the majority of students did *not* change their minds. In these cases, students noted that they gained reasoning skills or general knowledge about the case the most. Moreover, some students who gained perspective-taking argued that it helped their reasoning skills, too. For instance, one student said, "*I would say it helped justify my initial position because although I acknowledged their stances on the matter, I saw flaws in their reasoning and thus reinforced my opinion*." Interestingly, one student wrote, "*Discussing it with my groupmates reinforced my decisions, because I realized we simply have different values*." Similarly, another student wrote, "*I think discussing my opinion with others made me more aware of other perspectives. I wouldn't say it made me change my answer but I gave me more reasoning for why I chose my answer and why someone else would choose something different*." These responses indicated that realizing different moral perspectives helped the student accept that people have different core beliefs about things. Nevertheless, students of the following responses were typical and the most common type of response: "*I think that discussing with other people helped me justify my position even more because it made me argue and defend my position stronger*." Refer to Fig. 2 for a diagram that shows the relationship between students changing their minds and the aspects of discussions that were most interesting to them.

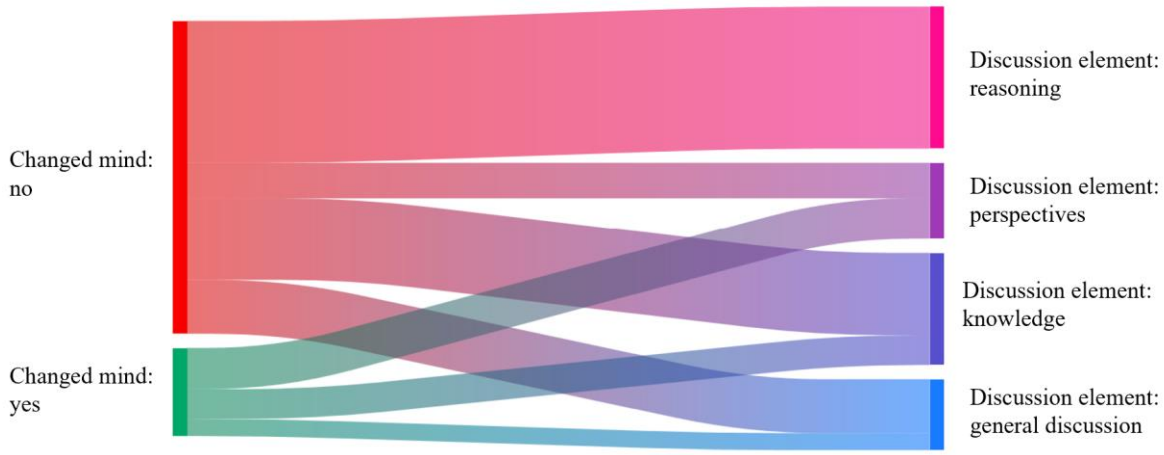

Fig. 2. Sankey diagram showing the relationship between whether students changed their mind after discussion and the most salient aspect of the discussion they held on to.

7. Discussion

In qualitative nature, we cannot infer whether students learn *more* when presented with ethical content designed by the SIM and DE principles than they would otherwise. But the results from this study showcase that students did respond positively to the novel approach to make intuitions front-and-center, and then use peer discourse to attempt to overcome the limitations of our gut-reaction judgment. Moreover, these results are aligned with evidence from moral psychology that students (a) are significantly impacted by perspective-taking [10], [13], (b) use reasoning to defend their position rather than seek and appraise truth [63], and (c) intentional focus on their morality lead to updated forms of knowledge [28].

In 2011, Zhong found that "knowing a course of action is wrong is not enough to prevent individuals from engaging in the action unless the action also 'feels' wrong…" [44, p. 18]. In the present study, measures were not used to assess students' *feeling*, but nevertheless, the data on whether students updated their decision may support Zhong's finding. Despite students gaining substantial reasoning skills and new knowledge about the ethical situation, the students did not change their minds. In fact, as seen in Fig. 2., the majority of the students who gained 'reasoning' and 'knowledge' from discussions did not change their minds. This may indicate that students need "something more" to change their minds (and thus change their course of action). This may be an update or change in how they pre-rationally *feel* about a situation rather than just what they *know* about a situation. Note, however, that changing one's mind is not necessarily good or bad; the dilemma was presented in a way that there was no correct answer, so a change of mind does not mean wrong to right or vice versa.

These results indicate that more work should be on connecting students' personal feelings of morality and how that connects to professional ethics standards that are expected of them. For instance, will students act on an ethical decision if they cannot "feel" the necessity of it? Additionally, more work should be done on the role of communication skills. Some students mentioned the ability to argue, defend, articulate, and communicate their ethical stance, but this learning intervention lacked the depth required to uncover how communication skills changed in response to the design of the learning intervention. This is argued by Neeley [25], as well, where the focus of communication in ethics education toward ethical action is often missed and overlooked. Additionally, from in-class researcher observations, some students were hesitant to enter discussions and were reserved while doing so. Future iterations of this learning intervention should consider the implications of developing a Brave Space before discussions to set an environment where students feel safe to enter morally contentious conversations [64], [65]. This may enhance the learning outcomes and better support communication skills, along with future recommendations to spend more time on the classroom-wide discussions.

8. Conclusion

This evidence-based practice paper introduced a learning intervention that synthesized recent findings from moral psychology and ethics pedagogy and qualitatively reported on students' perceptions of the learning intervention. In this, a learning design was detailed step-by-step, and pedagogical support for each step alongside the developed SIMDE conceptual framework. By qualitatively analyzing student reflections, this study found that students learned something new from this one-class-period intervention. Students were receptive to the design of the activity

itself; reflecting on the role of dilemma-based decision-making and transitioning from individual intuitive reflection to deliberative peer discourse was helpful. Moreover, the effects of students' peer discussions were in line with the literature, showing that even when students learn new information and improve their reasoning, they are not inclined to change their minds from their initial intuitive judgment. Also, this project has warranted more work from these authors that focuses on the distillation of this learning design to form an adaptable pedagogical framework based on the same founding principles of the SIMDE conceptual framework. As a result, this paper supports the need for updated applied ethics education in AI-adjacent curricula to respond to the quickly changing needs of future professionals.

9. Acknowledgments

This work was supported by the U.S. National Science Foundation under award # 2134667. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NSF or the U.S. Government.

References

[1] C. N. Bertolami, "Why Our Ethics Curricula Don't Work," *J. Dent. Educ.*, vol. 68, no. 4, pp. 414–425, 2004, doi: 10.1002/j.0022-0337.2004.68.4.tb03758.x.

[2] B. J. Grosz *et al.*, "Embedded EthiCS: integrating ethics across CS education," *Commun. ACM*, vol. 62, no. 8, pp. 54–61, Jul. 2019, doi: 10.1145/3330794.

[3] R. T. Javed *et al.*, "Get out of the BAG! Silos in AI Ethics Education: Unsupervised Topic Modeling Analysis of Global AI Curricula," *J. Artif. Intell. Res.*, vol. 73, pp. 933–965, Mar. 2022, doi: 10.1613/jair.1.13550.

[4] L. Tuovinen and A. Rohunen, "Teaching AI Ethics to Engineering Students: Reflections on Syllabus Design and Teaching Methods," 2021.

[5] J. Lönngren, "Exploring the discursive construction of ethics in an introductory engineering course," *J. Eng. Educ.*, vol. 110, no. 1, pp. 44–69, 2021, doi: 10.1002/jee.20367.

[6] R. F. Clancy, Q. Zhu, and Philosophy Documentation Center, "Why Should Ethical Behaviors Be the Ultimate Goal of Engineering Ethics Education?," *Bus. Prof. Ethics J.*, vol. 42, no. 1, pp. 33–53, 2023, doi: 10.5840/bpej202346136.

[7] N. Garrett, N. Beard, and C. Fiesler, "More Than 'If Time Allows': The Role of Ethics in AI Education," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York NY USA: ACM, Feb. 2020, pp. 272–278. doi: 10.1145/3375627.3375868.

[8] J. Southworth *et al.*, "Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy," *Comput. Educ. Artif. Intell.*, vol. 4, p. 100127, 2023, doi: 10.1016/j.caeai.2023.100127.

[9] C. Early, "A Case for Ethical Reasoning: Using the 8KQ to Guide Decision-Making in Daily Life," *Teach. Ethics*, vol. 22, no. 1, pp. 137–147, 2022, doi: 10.5840/tej2022221121.

[10] J. L. Hess, J. Beever, C. B. Zoltowski, L. Kisselburgh, and A. O. Brightman, "Enhancing engineering students' ethical reasoning: Situating reflexive principlism within the SIRA framework," *J. Eng. Educ.*, vol. 108, no. 1, pp. 82–102, 2019, doi: 10.1002/jee.20249.

[11] P. Murphy, "Teaching applied ethics to the righteous mind," *J. Moral Educ.*, vol. 43, no. 4, pp. 413–428, Oct. 2014, doi: 10.1080/03057240.2014.963036.

[12] E. L. Black, F. G. Burton, and J. K. Cieslewicz, "Improving Ethics: Extending the Theory of Planned Behavior to Include Moral Disengagement," *J. Bus. Ethics*, vol. 181, no. 4, pp. 945–978, Dec. 2022, doi: 10.1007/s10551-021-04896-z.

[13] J. Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychol. Rev.*, vol. 108, no. 4, pp. 814–834, 2001.

[14] L. Walker, "Gus in the Gap: Bridging the Judgment-Action Gap in Moral Functioning," in *Moral development, self, and identity*, D. K. Lapsley and D. Narváez, Eds., Mahwah, N.J: Lawrence Erlbaum Associates, 2004.

[15] E. Schwitzgebel and J. Rust, "The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior," *Philos. Psychol.*, vol. 27, no. 3, pp. 293–327, Jun. 2014, doi: 10.1080/09515089.2012.727135.

[16] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, "Conceptualizing AI literacy: An exploratory review," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100041, 2021, doi: 10.1016/j.caeai.2021.100041.

[17] Y. Zhou, Z. Zhan, L. Liu, J. Wan, S. Liu, and X. Zou, "International Prospects and Trends of Artificial Intelligence Education: A Content Analysis of Top-level AI Curriculum across Countries," in *Proceedings of the 6th International Conference on Digital Technology in*

Education, Hangzhou China: ACM, Sep. 2022, pp. 337–343. doi: 10.1145/3568739.3568796.
[18] C. Adams, P. Pente, G. Lemermeyer, and G. Rockwell, "Ethical principles for artificial intelligence in K-12 education," *Comput. Educ. Artif. Intell.*, vol. 4, p. 100131, Jan. 2023, doi: 10.1016/j.caeai.2023.100131.
[19] H. Bleher and M. Braun, "Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice," *Sci. Eng. Ethics*, vol. 29, no. 3, p. 21, May 2023, doi: 10.1007/s11948-023-00443-3.
[20] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, and M. Lennon, "Explaining the Principles to Practices Gap in AI," *IEEE Technol. Soc. Mag.*, vol. 40, no. 2, pp. 81–94, Jun. 2021, doi: 10.1109/MTS.2021.3056286.
[21] M. Davis, "Integrating ethics into technical courses: Micro-insertion," *Sci. Eng. Ethics*, vol. 12, no. 4, pp. 717–730, Dec. 2006, doi: 10.1007/s11948-006-0066-z.
[22] D. Elliott and K. June, "The Evolution of Ethics Education 1980–2015," in *Ethics Across the Curriculum—Pedagogical Perspectives*, E. E. Englehardt and M. S. Pritchard, Eds., Cham: Springer International Publishing, 2018, pp. 11–37. doi: 10.1007/978-3-319-78939-2_2.
[23] D. Bairaktarova and A. Woodcock, "Engineering Student's Ethical Awareness and Behavior: A New Motivational Model," *Sci. Eng. Ethics*, vol. 23, no. 4, pp. 1129–1157, Aug. 2017, doi: 10.1007/s11948-016-9814-x.
[24] A. Fox and B. C. Beiter, "Game Over: Reframing Ethical Decision-Making through Failure for Engineering Education," in *2023 ASEE Annual Conference & Exposition*, Baltimore, Maryland: ASEE Conferences, 2023.
[25] K. A. Neeley, "Knowledge Integration as the Foundation of Ethical Action: or, Why You Need All Three Legs of a Three-Legged Stool," in *2023 ASEE Annual Conference & Exposition*, Jun. 2023.
[26] J. Pijanowski, "Teaching Educational Leaders to Move from Moral Reasoning to Moral Action," *Educ. Leadersh. Rev.*, vol. 18, no. 1, pp. 37–51, 2017.
[27] S. B. Broomell and G. B. Chapman, "Looking Beyond Cognition for Risky Decision Making: COVID-19, the Environment, and Behavior," *J. Appl. Res. Mem. Cogn.*, vol. 10, no. 4, pp. 512–516, Dec. 2021, doi: 10.1016/j.jarmac.2021.10.003.
[28] L. Zollo, M. M. Pellegrini, and C. Ciappei, "What Sparks Ethical Decision Making? The Interplay Between Moral Intuition and Moral Reasoning: Lessons from the Scholastic Doctrine," *J. Bus. Ethics*, vol. 145, no. 4, pp. 681–700, Nov. 2017, doi: 10.1007/s10551-016-3221-8.
[29] R. J. Meisenbach, "Habermas's Discourse Ethics and Principle of Universalization as a Moral Framework for Organizational Communication," *Manag. Commun. Q.*, vol. 20, no. 1, pp. 39–62, Aug. 2006, doi: 10.1177/0893318906288277.
[30] J. Mingers and G. Walsham, "Toward Ethical Information Systems: The Contribution of Discourse Ethics," *MIS Q.*, vol. 34, no. 4, pp. 833–854, 2010, doi: 10.2307/25750707.
[31] J. S. Dryzek *et al.*, "The crisis of democracy and the science of deliberation," *Science*, vol. 363, no. 6432, pp. 1144–1146, Mar. 2019, doi: 10.1126/science.aaw2694.
[32] Santa Clara University, "A Framework for Ethical Decision Making." Accessed: Jan. 17, 2024. [Online]. Available: https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/

[33] L. Floridi and J. Cowls, “A Unified Framework of Five Principles for AI in Society,” *Harv. Data Sci. Rev.*, Jun. 2019, doi: 10.1162/99608f92.8cd550d1.
[34] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.
[35] *7 minutes to understand AI and key ROAM-X principles applying to its development*, (2021). [Online Video]. Available: https://www.youtube.com/playlist?list=PLWuYED1WVJIPHJLk84wWQbzeZcWLt5rwU
[36] P. Jenlink and K. Jenlink, “Ethical Leadership and Moral Literacy: Incorporating Ethical Dilemmas in a Case-Based Pedagogy,” *NCPEA Educ. Leadersh. Rev.*, vol. 16, no. 2, 2015, Accessed: Jan. 11, 2024. [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1105465.pdf
[37] B. Mittelstadt, “Principles alone cannot guarantee ethical AI,” *Nat. Mach. Intell.*, vol. 1, no. 11, pp. 501–507, Nov. 2019, doi: 10.1038/s42256-019-0114-4.
[38] L. Munn, “The uselessness of AI ethics,” *AI Ethics*, vol. 3, no. 3, pp. 869–877, Aug. 2023, doi: 10.1007/s43681-022-00209-w.
[39] M. D. Merrill, “First principles of instruction,” *Educ. Technol. Res. Dev.*, vol. 50, no. 3, pp. 43–59, 2002, doi: https://doi.org/10.1007/BF02505024.
[40] Z. Bagdasarov *et al.*, “Case-Based Ethics Instruction: The Influence of Contextual and Individual Factors in Case Content on Ethical Decision-Making,” *Sci. Eng. Ethics*, vol. 19, no. 3, pp. 1305–1322, Sep. 2013, doi: 10.1007/s11948-012-9414-3.
[41] J. F. Johnson *et al.*, “Case-Based Ethics Education: The Impact of Cause Complexity and Outcome Favorability on Ethicality,” *J. Empir. Res. Hum. Res. Ethics*, vol. 7, no. 3, pp. 63–77, Jul. 2012, doi: 10.1525/jer.2012.7.3.63.
[42] M. Khatiban, S. N. Falahan, R. Amini, A. Farahanchi, and A. Soltanian, “Lecture-based versus problem-based learning in ethics education among nursing students,” *Nurs. Ethics*, vol. 26, no. 6, pp. 1753–1764, Sep. 2019, doi: 10.1177/0969733018767246.
[43] M. H. Immordino-Yang and A. Damasio, “We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education,” *Mind Brain Educ.*, vol. 1, no. 1, pp. 3–10, 2007, doi: 10.1111/j.1751-228X.2007.00004.x.
[44] C.-B. Zhong, “The Ethical Dangers of Deliberative Decision Making,” *Adm. Sci. Q.*, vol. 56, no. 1, pp. 1–25, Mar. 2011, doi: 10.2189/asqu.2011.56.1.001.
[45] G. Gigerenzer, “Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality,” *Top. Cogn. Sci.*, vol. 2, no. 3, pp. 528–554, 2010, doi: 10.1111/j.1756-8765.2010.01094.x.
[46] B. Engelen, A. Thomas, A. Archer, and N. Van De Ven, “Exemplars and nudges: Combining two strategies for moral education,” *J. Moral Educ.*, vol. 47, no. 3, pp. 346–365, Jul. 2018, doi: 10.1080/03057240.2017.1396966.
[47] L. Kisselburgh, J. Hess, C. Zoltowski, J. Beever, and A. Brightman, “Assessing a Scaffolded, Interactive, and Reflective Analysis Framework for Developing Ethical Reasoning in Engineering Students,” in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana: ASEE Conferences, Jun. 2016, p. 26288. doi: 10.18260/p.26288.
[48] J. Beever and A. O. Brightman, “Reflexive Principlism as an Effective Approach for Developing Ethical Reasoning in Engineering,” *Sci. Eng. Ethics*, vol. 22, no. 1, pp. 275–291, Feb. 2016, doi: 10.1007/s11948-015-9633-5.

[49] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, and L. Floridi, “Ethics as a Service: A Pragmatic Operationalisation of AI Ethics,” *Minds Mach.*, vol. 31, no. 2, pp. 239–256, Jun. 2021, doi: 10.1007/s11023-021-09563-w.
[50] J. G. Finlayson and D. H. Rees, “Jürgen Habermas,” in *The Stanford Encyclopedia of Philosophy*, Fall 2023., E. N. Zalta and U. Nodelman, Eds., Metaphysics Research Lab, Stanford University, 2023. Accessed: Sep. 19, 2023. [Online]. Available: https://plato.stanford.edu/archives/fall2023/entries/habermas/
[51] G. Nunner-Winkler, “Discourse Ethics: A Pedagogical Policy for Promoting Democratic Virtues,” *Educ. Theory*, vol. 73, no. 2, pp. 273–291, 2023, doi: 10.1111/edth.12578.
[52] W. Rehg, “Discourse ethics for computer ethics: a heuristic for engaged dialogical reflection,” *Ethics Inf. Technol.*, vol. 17, no. 1, pp. 27–39, Mar. 2015, doi: 10.1007/s10676-014-9359-0.
[53] C. McIlwain, “AI has exacerbated racial bias in housing. Could it help eliminate it instead?,” MIT Technology Review. Accessed: Jan. 10, 2024. [Online]. Available: https://www.technologyreview.com/2020/10/20/1009452/ai-has-exacerbated-racial-bias-in-housing-could-it-help-eliminate-it-instead/
[54] R. Burns, “Artificial Intelligence Is Making The Housing Crisis Worse,” The Lever. Accessed: Jan. 10, 2024. [Online]. Available: https://www.levernews.com/artificial-intelligence-is-making-the-housing-crisis-worse/
[55] Géron, “Chapter 2. End-to-end Machine Learning Project,” in *Hands-on machine learning with Scikit-learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*, 2nd ed., Sebastopol: O’Reilly, 2019.
[56] J. Stefaniak, “Leveraging Failure-Based Learning to Support Decision-Making and Creative Risk in Instructional Design Pedagogy,” *TechTrends*, vol. 65, no. 4, pp. 646–652, Jul. 2021, doi: 10.1007/s11528-021-00608-6.
[57] D. G. Arce and M. C. Gentile, “Giving Voice to Values as a Leverage Point in Business Ethics Education,” *J. Bus. Ethics*, vol. 131, no. 3, pp. 535–542, Oct. 2015, doi: 10.1007/s10551-014-2470-7.
[58] B. Chang, “Reflection in Learning,” *Online Learn.*, vol. 23, no. 1, Mar. 2019, doi: 10.24059/olj.v23i1.1447.
[59] P. A. Ertmer and T. J. Newby, “The expert learner: Strategic, self-regulated, and reflective,” *Instr. Sci.*, vol. 24, no. 1, pp. 1–24, Jan. 1996, doi: 10.1007/BF00156001.
[60] ATLAS.ti, “ATLAS.ti Qualitative Data Analysis Software.” ATLAS.ti Scientific Software Development GmbH, 2023. Accessed: Jan. 21, 2024. [Online]. Available: https://atlasti.com
[61] M. Vaismoradi, J. Jones, H. Turunen, and S. Snelgrove, “Theme development in qualitative content analysis and thematic analysis,” *J. Nurs. Educ. Pract.*, vol. 6, no. 5, p. p100, Jan. 2016, doi: 10.5430/jnep.v6n5p100.
[62] M. Williams and T. Moser, “The Art of Coding and Thematic Exploration in Qualitative Research,” *Int. Manag. Rev.*, vol. 15, no. 1, pp. 45–55, 2019.
[63] J. Haidt, “The New Synthesis in Moral Psychology,” *Science*, vol. 316, no. 5827, pp. 998–1002, May 2007, doi: 10.1126/science.1137651.
[64] A. Cook-Sather, “Creating brave spaces within and through student-faculty pedagogical partnerships,” *Teach. Learn. Together High. Educ.*, vol. 1, no. 18, p. 1, 2016.
[65] B. V. Hole and M. De Luz, “An Imaginary of Radical Hope: Developing Brave Space for Class Discussion: Teaching Ethics,” *Teach. Ethics*, vol. 22, no. 1, pp. 83–96, 2022, doi: 10.5840/tej20221011119.