

“It’s OK Because...”: Students’ Rationalization of AI Use in Academic Writing

Jiyeon Kim¹, Kentaro Toyama², Sangmi Kim^{2,3}, John M. Carroll¹

¹Department of Information Science and Technology, The Pennsylvania State University

²School of Information, University of Michigan

³School of Education and Human Development, University of Miami

jxk6167@psu.edu, toyama@umich.edu, sxk2113@miami.edu, jmc56@psu.edu

Abstract

Generative AI challenges academic integrity not only by enabling students to delegate substantial portions of their academic work, but also by blurring the ethical boundaries between acceptable assistance and misconduct. Drawing on semi-structured interviews (n=20), AI chat logs, and course documents (syllabi, submitted assignments), we investigated how students themselves make moral sense of AI use in academic writing. Our analysis results in the following main novel findings: First, students use at least 23 distinct rationalizations to justify AI use, such as that copying AI-generated text is victimless; that any AI text reflecting their own beliefs or their own style is their own writing; or that they are learning more by using AI. We present a taxonomy of these rationalizations, comprising six core categories. Second, students’ rationalizations occur in both an ad hoc and post hoc manner and are sometimes lacking in normative grounding. These findings suggest that AI presents a slippery slope that students conceptually slide down, landing far outside the pedagogical goals and expectations of instructors. We discuss implications for educational design and AI policy.

Introduction

Nearly four years after the birth of generative AI (AI), significant debate over college students’ use of AI in higher education remains unsettled (Parker et al. 2026). This debate is particularly consequential for academic writing, which has long served as a central mode of learning through which students develop, organize, and demonstrate higher-order thinking (Emig 1977). Writing poses greater cognitive demands compared to reading or listening, as it requires a complex process involving the integration of multiple types of knowledge within the limited capacity of working memory (McCutchen 2000, 1996; Tindle and Longstaff 2015).

However, AI itself creates a crisis in academic writing and assessment, as students can have AI ghostwrite their work or easily humanize AI-generated text in final essays submitted for their courses (Wise et al. 2024; Giray et al. 2026). This poses significant challenges for instructors assessing students’ essays because it is hard to know which parts were written by the student and which by AI (Fleckenstein et al. 2024), while students are uneasy about not using AI for various reasons such as the perceived importance of the subject,

social norms, vague guidelines, and efficiency (Adnin et al. 2025; Giray et al. 2026).

Extensive research has examined how college students use AI in writing (Jelson et al. 2026), why students are hesitant to disclose their AI use when disclosure is required (Fu et al. 2026; Adnin et al. 2025), and how AI plagiarism/misconduct has emerged in higher education (Waqas, Hania, and Chunyan 2026; Bittle and El-Gayar 2025; Balalle and Pannilage 2025). Moreover, recent studies have investigated how existing moral frameworks predict students’ AI misconduct and cheating behavior (Theoharakis, Mylonopoulos, and Papadopoulou 2025; Hawdon, Costello, and Reichelmann 2026). However, little attention has been paid to why students believe the use of AI is morally acceptable, including cases involving noncompliance with course policy, and to how they justify such use in academic writing.

To understand students’ moral sense-making of AI use in academic writing, we conducted semi-structured interviews with 20 college students and collected AI chat logs (n=20), writing submissions, and course syllabi, where available. We contribute new insights into the following research questions:

- How do college students interpret and negotiate course AI policies in relation to their own beliefs and actual AI writing practices?
- How do college students justify their use of AI in academic writing?

Our results suggest that expectations surrounding AI use are distributed across five distinct contexts of AI-use conceptualization, ranging from formal policy to students’ actual AI use. Misalignment among these contexts creates openings for reinterpretation of AI use in practice. We identify a very broad range of rationalizations regarding harm, contribution, authorship, responsibility, and benefit — 23 distinct ones — that college students use to justify their use of AI in writing assignments. These are organized into six broader rationalization classes: Victimless Behavior (C1), Minimal AI Contribution (C2), Ex Ante Contribution (C3), Post Hoc Contribution (C4), Responsibility Denial (C5), and Perceived Benefit (C6). These rationalizations are often post hoc, inconsistent, and flexible, even when students themselves recognize conflicts with course policies or struggle with their own moral beliefs. This work contributes to meta-

level moral philosophy by examining the different ways college students think about the morality of AI use. We argue that AI policy and educational practice should move beyond articulating rules and instead help students understand the pedagogical rationale behind those policies.

Related Work

The Use of AI in Students' Writing

AI use among students is prevalent in academic contexts, including writing tasks. It was found to be very common that AI is involved in multiple phases of the writing process beyond just grammar checking (Bodei et al. 2026), including planning (asking for ideas, examples, or factual information), translating (turning student-provided ideas or partial text into prose), and/or reviewing (proofreading, feedback, or revision of user text) (Ammari et al. 2025; Jelson et al. 2026). Jelson et al. (2026) also found that students perform *vibe-writing*, a writing mode in which students guide ChatGPT through prompts while the AI generates most or all of the essay.

Moreover, AI has reshaped students' perceptions of ownership and authorship in writing. Participants were more likely to claim ownership of AI-generated text when asked to consider their roles in providing prompts and modifying the generated output (Wasi, Islam, and Islam 2024). Also, Jelson et al. (2026) found that students tend to feel a sense of ownership over their work when they use ChatGPT in the planning or reviewing stages of the writing process. Perception of contribution also varies by writing type. Participants reported less ownership of LLM-assisted creative writing, such as poems and stories, and attributed more of the contribution to the LLM than they did for non-creative writing, such as STEM assignments (Wasi, Islam, and Islam 2024).

AI Plagiarism and Misconduct in Student Writing

Historically, academic misconduct has been a major challenge (Eaton 2021). AI, however, has complicated conventional understandings of plagiarism. This is because AI-generated information may synthesize material from multiple sources, making its provenance difficult to trace (Huang and Chang 2024). Specifically, several types of AI-specific cheating behaviors have been reported in academic writing assignments. For example, students copy and paste AI-generated text without proper disclosure, a practice known as *AIgiarism* (Chan 2025). When using AI-powered writing tools, students often filter AI outputs through a tool that humanizes AI-generated text to conceal AI assistance. These forms of AI ghostwriting and hidden assistance pose new challenges for evaluating students' work properly (Scarfe et al. 2024; Giray et al. 2026), creating grey areas for academic integrity that fall beyond traditional plagiarism frameworks (Bittle and El-Gayar 2025; Balalle and Pannilage 2025) and raising broader questions of authorship, transparency, accountability, and learning assessment (Khlaif et al. 2025; Perkins 2023; Cotton, Cotton, and Shipway 2024; Wise et al. 2024; Khosrowi, Finn, and Clark 2023).

In response, institutions have developed AI governance frameworks, revised academic integrity policies, and re-

designed assessments to address AI-enabled misconduct (Wu, Zhang, and Carroll 2024; Khlaif et al. 2025). At the course level, instructors have adopted assessment strategies that emphasize process over output (Slimi 2026) and have used AI-detection tools. However, these tools remain unreliable, raising concerns about bias and inequity and increasing student anxiety by falsely flagging students who have not used AI (Ardito 2025; Liang et al. 2023; Slimi 2026). Instructors have also required students to disclose their use of AI, yet students often hesitate to do so (Gonsalves 2025; Adnin et al. 2025). Rather than treating AI use as inherently unethical, Perkins (2023) argues that its ethical status depends on transparent use and alignment with institutional expectations.

Some scholars argue that it is important to reconsider academic integrity both in principle and in practice when AI could replace almost all human writing processes (Eaton 2023; Chan 2023; Slimi 2026). Rather than arguing for banning AI, researchers have suggested reconfiguring academic environments to support learning and to establish legitimate authorship and responsibility for AI use (Furze et al. 2024; Foltyniek et al. 2023). Yet, there is still no clear consensus on what constitutes AI cheating (Gruenhagen et al. 2024; Perkins and Roe 2024). Some studies define it primarily through authorship, treating AI cheating as submitting AI-generated work as one's own (Chan 2025). Others frame misconduct as the use of AI assistance without permission, acknowledgment, or disclosure, or in ways that conflict with institutional expectations (Foltyniek et al. 2023; Perkins 2023). Assessment-focused approaches further show that AI use may be acceptable or unacceptable depending on task-specific rules and permitted levels of AI assistance (Perkins et al. 2024; Perkins and Roe 2024).

Furthermore, there is a gap between students and instructors regarding the extent to which students may use AI and incorporate AI-generated work into academic practices, as well as what constitutes an ethical standard for such use. For example, Lin et al. (2026) found broad agreement on using AI for ideation and revision but substantial disagreement over AI-generated text. Instructors strongly opposed the use of such text, whereas students were more willing to incorporate it into their work.

AI Use and Moral Framework

Ethical judgments about academic integrity can be understood through several normative views. From a deontological perspective, misconduct concerns whether students violate rules or duties (Alexander and Moore 2007); from a virtue-ethics perspective, whether their behavior conflicts with being an honest or responsible student (Hursthouse and Pettigrove 2000); and from a consequentialist perspective, whether the perceived benefits of an action outweigh its harms/risks (Scheffler 1988). These lenses help understand whether students view AI use as breaking the rules or a trade-off among learning, efficiency, and potential consequences.

Recent studies have applied existing moral frameworks to predict AI misconduct from students' moral reasoning and affect. For example, Theoharakis, Mylonopoulos, and

Papadopoulou (2025) found that moral disengagement predicted misconduct among UK graduate business students, with stronger effects when students perceived AI as useful, used it habitually, or had stronger prompting skills. Similarly, Hawdon, Costello, and Reichelmann (2026) found that neutralizing excuses were associated with AI cheating, whereas anticipated guilt was negatively associated with it.

Recent studies have begun to examine how students justify their use of AI. For example, using the Theory of Planned Behavior (TPB) as a framework, Giray et al. (2026) interviewed six Filipino undergraduates (who admitted to cheating on writing assignments) and found that participants reframed this behavior as a legitimate academic survival strategy rather than as misconduct. They identified the perceived relevance of courses to students' academic and career goals, peer normalization, institutional ambiguity, and confidence in avoiding detection.

While prior studies have examined how established moral frameworks relate to students' AI misconduct or how students admit to cheating while justifying their use of AI in writing, less is known about how students themselves define the boundary of ethically acceptable AI use in academic writing.

Method

Participants and Recruitment

To answer our RQs, participant recruitment targeted college students who self-reported AI use for writing assignments. Participants were recruited through poster flyers in campus buildings, an institutional platform for study participant recruitment (e.g., StudyFinder), and a social media post (on Discord). Those who volunteered to participate in our study were instructed to complete a screening survey via Qualtrics that asked how often they used AI for writing assignments on a 5-point Likert scale: never, rarely, sometimes, often, and always. Based on the screening survey responses, we invited those who selected "sometimes" or higher to the individual interviews. Or, we manually reviewed their submitted prompts to determine whether their actual AI use was substantial enough for inclusion. The final sample consisted of 20 undergraduate students from 12 different U.S. universities (15 females and 5 males). 14 of them were domestic students, 5 were international students, and 1 was a student with dual nationality. Participants' majors were diverse, including accounting, pre-med, communication, statistics, psychology, finance, nursing, education, English, art, or cell and molecular biology. All interviews took place between December 2025 and May 2026, and participants were compensated with a \$20 Amazon gift card.

Data Collection

We collected interview data and AI prompts/chat logs from all participants ($n=20$), as well as course syllabi and writing assignments when available. Before participating in individual interviews, participants were asked to share their AI prompts and chat logs used to complete writing assignments. We also asked students to share the course policy on AI from the course syllabus or the instructor's in/formal instructions,

and the final essay submissions from the course. The submission of this information was completely voluntary; 5 students shared their final essays, 9 shared an AI policy for the course as part of the syllabus, and 7 shared the instructor's in/formal instructions. Each participant shared chat logs via Qualtrics¹. The submitted information was used to understand participants' AI use history and better craft questions relevant to each participant before the interview sessions. It was also reviewed with the participant during the interview and used to elicit more honest and vivid responses.

Semi-structured interviews were conducted virtually and lasted approximately one hour. As AI use is a sensitive topic and students are commonly hesitant to share their use (Adnin et al. 2025), participants were informed that all information shared during the interview would be confidential and were encouraged to share their honest experiences and thoughts about AI use in writing assignments. Each interview began with questions about what the assignment was about, what their prompts aimed to do, why they asked AI such questions, and what each participant expected AI to answer. Then, follow-up questions were asked to understand their awareness of the current AI policy within the course and their thoughts about that policy. Finally, participants were asked to share their honest thoughts on their current use of AI regarding learning experiences and ethics. This study was approved by the IRB at Pennsylvania State University.

Data Analysis

Due to IRB restrictions on data sharing (across other authors from different institutions), only the first author and last author had access to the students' raw chat logs. Following discussion with and advice from the IRB liaison, the first author de-identified all participants' personal information, such as names, email addresses, and course names, before sharing the students' chat logs, or interview transcripts with other research team members, replacing identifiers with a Participant ID (PX). This constraint shaped our analytic workflow. The first author conducted hands-on coding of raw data, while the research team participated in interpretation, refinement, and consensus-building over de-identified data.

We conducted thematic analysis, allowing codes and themes to emerge from the data without a predetermined coding framework (Clarke and Braun 2017). First, the first author reviewed interview recordings and transcripts multiple times and identified portions of the interview transcripts in which participants described or justified their use of AI in academic writing. Using these excerpts as units of analysis, we conducted several rounds of iterative thematic clustering, beginning with an initial organization of participants' utterances around recurring forms of ethical reasoning. As we compared excerpts across participants, it became clear that participants were not simply describing whether they used AI, but repeatedly offering different moral justifications for why particular AI-assisted writing practices were

¹Participants were asked to share one full AI-assisted writing session, defined as a continuous conversation with an AI tool from the time they began working on an essay assignment until they stopped working during that session.

acceptable, necessary, harmless, or ambiguous. We therefore focused our analysis on the rationalizations participants used to make sense of their AI use, rather than on classifying AI-use practices alone.

Through this process, we initially clustered individual excerpts into 46 subcodes that captured recurring patterns in participants' rationalizations. These subcodes were then consolidated into 23 rationalization categories based on shared similarities, which were further grouped into six broader rationalization classes: victimless behavior, minimal AI contribution, ex ante contribution, post hoc contribution, responsibility denial, and perceived benefit (see Table 1 and Table 2). Definitions of each rationalization were developed inductively from participant accounts and refined through discussion among all research team members. Due to space constraints, we report the 23 category level with six classes in this paper.

At that point, data were analyzed in two complementary ways. First, iterative thematic clustering continued in order to identify higher-level patterns across participants' moral reasoning. Second, excerpts that reflected ethical reasoning about AI use were coded according to rationalization categories wherever sufficient information was available. Throughout the process, the research team met on a regular basis to compare interpretations, discuss similarities and differences across rationalizations, refine category boundaries, and resolve disagreements through consensus. This process resulted in 23 student rationalizations as well as the broader findings regarding students' flexible, post hoc, and sometimes internally inconsistent moral sense-making around AI use in academic writing. The research team reviewed all the codes, and whenever there was disagreement with the first author (<5% of instances), the authors discussed to arrive at a consensus.

Findings

Five Contexts of AI "Policy"

Our analysis of syllabi and interviews together suggests that for any given course, there are five distinct "contexts" that house concepts of appropriate AI use for academic writing: (1) **Faculty Intention**, instructor expectations of student AI use; (2) **Formal Policy**, the explicitly specified rules, usually written down in syllabi or assignment instructions; (3) **Student Interpretation**, a student's own understanding, recall, and/or interpretation of formal policy; (4) **Students' Own Beliefs**, about roles of AI in writing assignments; and (5) **Student Practice**, a student's actual use of AI.

We find that the AI "policies" – whether explicit or implicit, formal or informal, specific or vague – that are expressed in each of these contexts often differ from the others (though, not necessarily). For example, while we did not interview faculty directly, it was clear from syllabi and participant interviews, that Faculty Intention (Context 1) was not always clearly expressed in their course's Formal Policy (Context 2). One participant (P8) explained that they could copy and paste AI-generated text because the course syllabus allowed AI use for idea generation. Whether the instructor intended to permit direct copying from AI output

was unspecified in the syllabus, though presumably, they had some intention about it. Relatedly, a faculty co-author of this paper notes that in their own teaching, it took several iterations of a course to converge on a well-specified written policy that expressed their expectations of AI use – in other words, their first attempts did not perfectly match their intention.

Similarly, some students misremembered their course AI policies, or mixed up policies between courses, demonstrating the distinction between Formal Policy (Context 2) and Student Interpretation (Context 3). Some of our participants expressed that they knew what the course policy was Student Interpretation (Context 3), but applied a policy that made more sense to them Students' Own Beliefs (Context 4). For example, some participants suggested that strict AI prohibitions were unrealistic or educationally unhelpful, even when they recognized that instructors might disapprove of their use. And finally, we saw several instances where students violated their own stated norms Students' Own Beliefs (Context 4) in actual practice Student Practice (Context 5). In some cases, the boundary of acceptable AI use was relocated across multiple contexts. One participant (P11), for example, recalled that the course's formal policy prohibited the use of AI and that the instructor justified this restriction by explaining that "writing is thinking" and that students should "not use AI to think for you." Nevertheless, P11 gave AI their ideas, asked it to expand on them, and copied and pasted the resulting text. P11 justified this practice by arguing that AI had helped refine their thinking rather than think on their behalf.

It might seem obvious that AI use is shaped not only by faculty expectations and formal rules, but also by students' interpretations, personal beliefs, and ultimate practice. Yet, the fact of these five contexts suggests that AI presents myriad opportunities for misinterpretation, second-guessing, confusion, and departures from course AI policies.

The Diversity of Moral Justifications for AI Use

We identified 23 distinct moral justifications that participants mentioned regarding their AI use – henceforth "Rationalizations" (R1–R23). These 23 Rationalizations can be grouped into six higher-level categories, which we call *Rationalization Classes* or simply, *Classes* (C1–C6). Table 1 and Table 2 present our taxonomy, along with representative participant quotes for each Rationalization. Space does not permit an in-depth description of every Rationalization, but below, we describe each Rationalization Class and offer examples of the Rationalizations that it comprises².

The first Class, **C1: Victimless Behavior**, involves two Rationalizations in which participants suggest their AI use is acceptable, because no (human) person is harmed by it,

²Participants' use of AI in writing practices varied, ranging from making an outline, paraphrasing AI-generated text, and expanding self-written texts to copy-pasting AI-generated text. Our taxonomy classifies the rationalizations participants used to justify these practices, rather than the practices themselves. Thus, the same practice could be justified through different rationalizations, and the same rationalization could support different levels of AI involvement.

or because the only person being harmed was the participant themselves. Some participants implied that plagiarism was bad only because it hurt others, but as P16 noted, “*AI doesn’t have a spirit and a soul... giving AI credit... and respect as if AI had ownership or legality to certain information... that’s simply not true.*” A few participants stated directly that copying and pasting AI-generated text did not harm other human beings and therefore did not raise an ethical problem. Similarly, some participants argued that AI synthesizes information rather than copying from a specific source, thus eliminating any human victims of plagiarism. Interestingly, one participant noted that there was no harm to instructors, and went as far as to suggest the only possible harm as being to oneself: “*It’s not hurting anybody ... I think you are hurting yourself more than you’re hurting like a teacher or other people*” (P6).

The second Class, **C2: Minimal AI Contribution**, involves Rationalizations where AI’s contribution is cast as trivial or equivalent to other allowable forms of assistance, thereby limiting its moral significance. This Class includes *R3: Busywork*, the most common Rationalization raised by participants. They argued that AI was OK for “busywork,” because such assignments were devoid of learning value, unimportant, meaningless, or low-stakes (unlike, e.g., exams). A related Rationalization was *R5: Like Other Allowable Support*, where participants claimed that AI was similar to traditional tools or forms of support, such as human editors, calculators, grammar checkers, or writing centers, and therefore also acceptable to use. For example, one participant compared AI to receiving suggestions from a human writing coach, even though the text that expanded the idea was generated by ChatGPT: “*If I don’t have that answer for myself, then she [the coach] would say, you could do this. It’s like, the same as [what] ChatGPT is doing for me.*” (P14). Similarly, one participant compared using AI-generated text to using a calculator’s answers, which is not considered copying or taking someone else’s idea: “*If you ask a calculator to give you like do something for you that you’re not exactly copying it, right? It’s just something that a machine is giving you.*” (P11).

The common logic across this Rationalization Class was that AI’s contribution to the final work was not substantial enough to threaten authorship or academic integrity. Rather than denying that AI had contributed, participants reframed its use as minor assistance.

The third Class, **C3: Ex Ante Contribution**, involves Rationalizations in which participants provided some core content before turning to AI and emphasized that this prior input was the critical part of the work. These rationalizations suggest that AI use is acceptable because the core idea, task direction, or source materials came from the participant. For example, Rationalization *R6: My Ideas* involves cases in which participants described already having ideas in mind but needing help to articulate or to expand on them. One participant explained that they had “*Ideas in my mind, but it’s hard to articulate*” (P3), while another said, “*I have the story in my mind, but I really wanted [AI] to expand it for me*” (P14). Similarly, Rationalization *R7: My Directions* refers to cases in which participants explained that their pro-

vision of detailed prompts to AI was the essential work.

With Class C3, the student sees themselves as the primary originator or director of the work, and therefore creatively and ethically above board in using AI. Like the previous Class C2, Minimal AI Contribution, this Class also discounts AI’s role, but here, the Rationalizations prioritize the student’s substantive and ideational contributions as primary and essential.

In contrast to C3, the fourth Class, **C4: Post Hoc Contribution**, involves Rationalizations in which the student’s contribution occurs after AI has output substantive content, and often only in stylistic or confirmatory ways. Thus, with Rationalization *R10: My Paraphrasing*, students argue that paraphrasing AI output makes it their own; or, with Rationalization *R11: My Verification*, that checking AI output for accuracy is sufficient and proves responsible use of AI.

The most interesting in this Class is Rationalization *R12: My Style*, in which participants explain that if the output “sounds like” them, it is as good as if they wrote it. One participant thinks it is more ethical to use AI-generated text when “*It’s aligning with my personal beliefs.*” (P7). A few participants described either prompting or experiencing AI take on their writing style. One participant explained that they could even ask AI to generate prompts that imitated their thinking style:

“It has recognized a pattern of my way of writing and how I express things in, like projects or emails or anything like that. So now, it knows what kind of person I am like in the way of writing. So now, it gives the prompts similar to like, what my thinking process is” (P9).

The fifth Class, **C5: Responsibility Denial**, comprises Rationalizations that weaken, displace, or externalize participants’ sense of responsibility. With Rationalization *R13: Normalization*, participants pointed to external norms, arguing that AI use had become socially normalized and practically unavoidable. Several explained that they needed to use AI to keep up and avoid being placed at a disadvantage: “*If I am the only one writing my own work and everyone else is using ChatGPT, I feel like they’re ahead of me because they just saved four hours.*” (P18) Some participants also raised fairness concerns, especially when classmates appeared to use AI and received good grades: “*Some of my classmates who just throw it [into AI], generate it, don’t even check it, [and] they get higher grades than I do*” (P17). Participants also invoked *R15: Instructor Indifference* and *R17: No Consequences* as justifications. For example, some participants blamed instructors, arguing that assignment instructions were ambiguous or difficult to understand:

“Some professors won’t really help you with, even getting to this train of thought. So I think you can use this[AI] because if professors aren’t helping you at all, you need to use something to help you learn.” (P8)

Others emphasized the absence of consequences, suggesting that AI use felt acceptable because it was unlikely to be detected or punished. As with other Rationalizations in Class

C5, participants shifted responsibility away from themselves or denied any wrongful intention.

The sixth and last Class, **C6: Perceived Benefit**, justifies AI use through its practical, educational, or performance-related advantages. Many participants, for instance, invoked Rationalization *R19: Time Economy*, stressing that AI use saved them time. One participant went as far as to say, “[I would rather] spend time with my girlfriend...” (P13). Notably, this Rationalization was frequently paired with Rationalization *R3: Busywork*.

Some participants believed that they could still learn new material by reading AI-generated text and later apply that knowledge when completing similar tasks independently, while others viewed paraphrasing AI-generated text as a way to preserve learning, reasoning that they still learned by rewriting the content in their own words (*R20: Educational Value*). Others perceived AI-generated writing as higher quality than their own writing and argued that this could support better learning (*R21: Better Writing*). One unique participant (P6) claimed that AI use was ethically acceptable if the user had some degree of sincere intention to learn (*R22: Learning-oriented Intention*). Participants also justified AI use by emphasizing positive outcomes or performance benefits, such as receiving a good grade, earning faculty praise, or simply completing the assignment (*R23: Better Outcome*). In all of the Rationalizations in Class C6, participants reason that AI use’s beneficial effects make it permissible – the ends justified the means.

Belief vs Behavior Mismatch

We have noted already that participants’ ideas about appropriate AI use, whether based on Student Interpretation of policy or Students’ Own Beliefs, can diverge from actual Student Practice.

A few participants continued using AI despite recognizing that doing so violated course rules or conflicted with their moral commitments because the faculty policy did not convince them. P16, for example, expressed a strong desire to follow faculty rules and act ethically, stating, “I want to do what my teacher is asking me to do. I want to obey her, and I want to do what’s right” (P16). Yet they also described continuing to use AI despite a strict course policy prohibiting it and expressed strong frustration with the AI policy. (e.g., “[The instructor says] never to use AI. Ever, never, ever, ever, ever, ever. She’s like, don’t look at it, don’t smell it, don’t touch it” (P16)). Because an outright prohibition is impossible to follow as AI has become a way of life for their generation, P16 wants faculty to develop AI policies that students who genuinely intend to act ethically can realistically comply with:

“Instead of being delusional... you should [try] to encourage AI policies that students [who] have a desire to be morally right can actually walk out [into] the real world.” (P16)

Other participants acknowledged that some of their practices were “not very ethical” (P15) but treated such concerns as secondary to immediate pressures and practical considerations, such as time constraints, workload, convenience, and

other priorities. For example, P15 explained that they had to focus on a major exam and therefore “put aside [their] ethical concerns” (P15). Similarly, P19 framed the decision as involving a trade-off between morality and time:

“In this paper, I didn’t follow my moral compass, and at the end, I just copy-pasted the conclusion... It depends how much I’m willing to balance my morals. There’s like balancing my time as well.” (P19)

A distinct pattern involved managing the visibility of AI use rather than questioning its morality. In these cases, participants did not regard their AI use as morally problematic, yet concealed it from faculty members or attempted to make it less visible. For example, some participants considered paraphrasing AI-generated text acceptable but still ran their drafts through AI detectors before submission to reduce the likelihood that their AI use would be detected.

A further discrepancy emerged between participants’ stated ethical boundaries and their actual practices. Several maintained that AI use was ethically acceptable as long as they did not copy and paste AI-generated text. However, comparisons between their final essays and AI chat logs indicated that they had, in fact, copied and pasted such text. Three participants did not recognize this discrepancy until they reviewed their final submissions during the interviews.

Unfinished Rationalizations

Thus far, we have analyzed participant rationalizations of their AI use in isolation. However, our interview transcripts also exhibit tortuous routes of moral reasoning in which rationalizations evolved during the interview; rarely did students start with a single rationale and remain with it. Their reasoning was fluid, unfinished, and sometimes internally contradictory. Among the 20 participants, 16 exhibited this pattern.

Given space constraints, we use P6 as a particularly vivid example of the back-and-forth contradictions observed across participants. P6’s reasoning moved among *R3 Busywork*, *R8 My Curation*, *R10 My Paraphrasing*, *R11 My Verification*, *R12 My Style*, *R19 Time Economy*, *R20 Educational Value*, and *R22 Learning-oriented Intention*.

P6 example. P6 showed unresolved moral ambivalence through shifts in how they evaluated AI use. P6 admitted that completing homework with AI, including copy-pasting AI-generated text, could be unethical and reduce learning, yet often did not disclose their AI use because it was “Really hard for [faculty] to prove.”

“Honestly, it’s probably not very ethical. I do a lot of my homework with AI, and it doesn’t always mean I’m learning.” (P6)

This uneasiness coexisted with an appreciation of AI’s time-saving benefits. P6 distinguished between more and less ethical uses of AI based on the task and the degree of personal involvement. AI use in writing assignments was perceived as more ethical because P6 provided the information to be incorporated (*R8 My Curation*), used AI-generated text that was “Still words I agree with and I believe in” (*R12 My Style*), and reviewed the text for accuracy (*R11 My Verification*).

#	Rationalization	Definition	Representative Quote
Rationalization Class C1: Victimless Behavior — “AI use has no human victim; ethics are inapplicable.”			
R1	No Human Victim	It is OK because plagiarism and authorship norms exist to protect human authors, human effort, and human ownership. Since AI is not a person, does not exert human-like effort, and no human is directly harmed, AI-generated text lacks a morally relevant victim.	“It wasn’t written by another human... it wasn’t somebody’s idea that I stole... So I guess it’s fine” (P11)
R2	AI-Synthesized Sources	It is OK because AI synthesizes and rewrites information rather than copying from a specific source; therefore, I do not see it as plagiarism.	“With ChatGPT, [it] is not copy pasting from an external source. What it’s doing is analyzing the information from four or five different sources, then clubbing it into one, writing it as a whole different thing... So this is not plagiarism, because this is not directly copy paste [from] the sources” (P9)
Rationalization Class C2: Minimal AI Contribution — “AI contribution is too minimal to be seen as an ethical issue.”			
R3	Busywork	It is OK because AI contribution to only “busywork.”	“With a lot of these assignments, they’re not particularly... challenging, but it’s just a lot of busy work... it’s just not worth my time to actually... consume all of that content” (P2)
R4	Facts Only	It is OK because factual information cannot be owned.	“When [assignment] asks you to write about the facts... you can use ChatGPT, because the facts are not going to change” (P5)
R5	Like Other Allowable Support	It is OK because AI’s role is the same as existing writing resources and tools such as the writing center, proofreading editor, Grammarly, Calculator, and Google.	“The writing center does the same thing... she’s also... asking me to think, but it also helping me, like a booster... It’s like, the same as the ChatGPT is doing it for me” (P14)
Rationalization Class C3: Ex Ante Contribution — “I take care of the critical parts and exert control over the AI processes and outputs.”			
R6	My Ideas	It is OK because the core ideas, intentions, or thoughts originate from me. AI helps articulate, expand, or clarify what I already had in mind.	“When I get a topic in my head, at that point, I know what the topic was. I know what it needs... but at that point it’s like jumbled up in my head. I can’t really put out... an organized statement... So [AI] organized my thoughts in a better way” (P10)
R7	My Directions	It is OK because I direct AI through prompts.	“I’m the one who gave it a prompt. So I already dictated for it what it should give me, what I need... and how I want the information to be dispensed back to me” (P7)
R8	My Curation	It is OK because I do the research or select relevant materials before using AI. I select what information matters before AI performs the writing or organizing task.	“I did most of the work, especially because I put in, like, a filled-out outline with all the sources and stuff. And all [AI] really did was reword the way the sources wrote. It had to be in essay format” (P12)
Rationalization Class C4: Post Hoc Contribution — “I revise or check AI’s work afterward.”			
R9	Selective Use	It is OK because I do not take AI text wholesale. I take only portions of it.	“I guess that would be fine. Like, one or two sentences [copy-pasted] is fine. I guess it’s not like blatant copying” (P11)
R10	My Paraphrasing	It is OK because I transform AI output through editing, addition, or iterative refinements.	“I basically just paraphrase it and kind of condense it... I just go sentence by sentence, like condensing and paraphrasing each sentence in my own words for the assignment” (P2)

Table 1: Taxonomy of Student Rationalizations for AI Use in Academic Writing.

#	Rationalization	Definition	Representative Quote
<i>Rationalization Class C4: Post Hoc Contribution (continued)</i>			
R11	My Verification	It is OK because I verify, fact-check, and cite the original sources.	"I'm actually going back to like, check it and make sure everything is like, correct" (P17)
R12	My Style	It is OK to use AI if it sounds like me.	"I have spent a lot of time training my ChatGPT to sound like me and like to give good responses, so whenever [I] put in [a prompt], I'm confident that it's going to be sounding like me" (P9)
<i>Rationalization Class C5: Responsibility Denial</i> — "There is nothing wrong with me using AI because it has now become common and no one cares."			
R13	Normalization	It is OK because everyone uses AI. It is normal.	"I think a big thing for me that's not a deterrent, is everyone uses it..." (P3)
R14	Inevitability	It is OK because AI use is inevitable.	"We already have to accept this is a technology that is part of us, and as time goes by, it is continuing to develop, and generally, those who don't use it will be left behind" (P13)
R15	Instructor Indifference	It is OK because faculty do not care, give vague instructions, or implicitly permit AI use.	"They [faculty] don't really care about [assignments]... they don't even look at it really... they never write notes... never give us feedback. Again, they're only looking at our exams" (P18)
R16	Normlessness	It is OK because there are no clear rules about AI use yet.	"AI is so Wild Wild West... there's not a lot of rules surrounding it... since it's such a new technology, we haven't really grasped the concept" (P19)
R17	No Consequences	It is OK because there are no consequences, such as unlikely to be detected, punished, or grade penalized.	"It feels like a guilty conscience, but I'm still submitting it because I know that I'm not going to get in trouble" (P15)
R18	Agency Denial	It is OK because the action is not really mine; either AI performed it, or I did not consciously or intentionally act.	"If [there is] anything AI plagiarizes, AI is the one plagiarizing, because it's getting information from the Internet" (P16)
<i>Rationalization Class C6: Perceived Benefit</i> — "Using AI is beneficial for me."			
R19	Time Economy	It is OK because of time and effort considerations ranging from convenience to necessity, such as overwhelming workload or language barriers.	"Instead of spending four hours filling out this paperwork, it kind of takes 30 minutes" (P18)
R20	Educational Value	It is OK because I still learn the content through AI use.	"When I put into AI summary, like while I'm [paraphrasing AI's essay sentence by sentence], I still learn the content. Like while I'm writing it, since I am typing it out, the information from the documentary and the information from the book that I should have read about is still going into my brain because I'm reading this stuff from the AI [essay], and I'm [re]writing the essay" (P2)
R21	Better Writing	It is OK because AI is a better writer than me.	"I feel like I'm more confident when ChatGPT writes it for me because it has better approaches to grammar and transitioning phrases and structuring" (P9)
R22	Learning-oriented Intention	It is OK when my intention is to learn.	"It's a lot about the intentions that you have when you're using it. If you're trying to just get the class done and you don't really care, then obviously you're hurting yourself and learning less, and that's more unethical... But if you are trying to get something out of it... then it's more ethical" (P6)
R23	Better Outcome	It is OK because the outcome is good, such as a better grade or faculty approval.	"Before AI, my essays were terrible... now it's easier to actually go for an A than just go for a C" (P12)

Table 2: Taxonomy of Student Rationalizations for AI Use in Academic Writing.

P6 similarly explained that *“More of it will end up being my original work by the end, because I will change it”* (R10 *My Paraphrasing*). This distinction centered on whether P6 had added their own thoughts. In contrast, P6 described submitting an AI-generated discussion post without reading the textbook as *“Probably less ethical because I didn’t read the textbook... I didn’t really include any of my own thoughts.”*

P6 initially framed the harms as primarily personal, arguing that using AI-generated text undermined their own learning more than it harmed others. But later, P6 recognized broader instructional harm: if instructors cannot distinguish students’ work from AI-generated content, they cannot accurately assess learning or provide appropriate feedback. Yet P6 softened this assessment, arguing that reviewing AI-generated text reduced the harm to their own learning because they still learned something in the process (R20 *Educational Value*).

P6 also acknowledged that not following course AI policies, including failing to disclose or cite AI use, was unethical: *“If they just say specifically not to use AI and I do it, and I don’t cite it, I think that’s definitely unethical, even though I still do it.”* They added, *“If I have a teacher that doesn’t want me to use AI at all, I’m still gonna do it, but I’m lying to them now, and that’s wrong.”*

Yet P6 shifted again by minimizing the importance of certain classes (R3 *Busywork*): *“I use AI a lot more in a class that I just want to pass, and I don’t really care about it.”* P6 also reframed AI use through intention, suggesting it was more ethical when the goal was to save time or gain some value rather than simply avoid work (R19 *Time Economy*, R22 *Learning-oriented Intention*): *“For me, it’s a lot about the intentions that you have when you’re using it.”*

Finally, P6 worried that growing dependence on AI was eroding their ability to articulate thoughts in real-life conversations: *“I think that I’m losing some of the skills to talk to other people... it’s hurt me.”* Yet P6 immediately explained why they continued:

“But again, when I use it in the short term, I don’t think too much about that long term effect. I just think about, I just want this assignment done, and that’s it.”
(P6)

One can infer from the above that college students’ rationalizations about AI use are often post hoc. Participants often contradicted themselves across their multiple rationalizations. Some participants acknowledged that they had not thought about the ethics or morality of their AI use until interview questions brought them to mind: *“Over time, we just do it so much that we don’t really think about if it’s a good idea or not.”* (P6)

Discussion

Our findings demonstrate 23 distinct rationalizations that university students use to justify their use of AI in writing. Students interpret acceptable AI use in terms of harm, authorship, responsibility, contribution, and benefits. Students’ rationalizations were often used to explain behaviors that had already occurred, rather than to plan future behavior.

These post hoc rationalizations have been reported in the literature on motivated moral reasoning (Ditto, Pizarro, and Tannenbaum 2009) and the social intuitionist model (Haidt 2001), implying this type of justification may lack normative grounding.

Unlike familiar forms of academic misconduct, which are often governed by well-established norms and relatively clear ethical standards (Maramba 1993; Stearns 1992), AI-assisted writing presents students with a new and normatively unsettled situation. Students are therefore compelled to work out for themselves how harm, authorship, responsibility, and contribution should be understood. In traditional plagiarism, the prohibited act, the original author, and the misrepresentation of authorship are relatively identifiable. For example, traditional plagiarism typically involves an identifiable victim (the original author) whose authorship is denied (Stearns 1992). With AI, however, students may argue that no human victim exists and that, therefore, no ethical problem is involved, as found in **C1: Victimless Behavior**. Also, **C2: Minimal AI Contribution**, **C3: Ex Ante Contribution**, and **C4: Post Hoc Contribution** show that students minimize AI’s role while emphasizing the significance of their own contributions – such as originating ideas, prompting AI, editing its output, and verifying AI-generated text – to justify treating the resulting work as their own and their use of AI as ethically acceptable. At the same time, while students consider AI output their own, they often distance themselves when it can create potential risks like plagiarism or policy violations, as in **C5: Responsibility Denial**, thereby externalizing responsibility in ways that echo moral disengagement (Bandura 2011). Together, these different and at times conflicting rationalizations show that making sense of the authorship of AI-generated outputs and of responsibility for AI use is contingent on context.

AI-assisted writing challenges not only conventional understandings of academic misconduct but also the basic moral principles of authorship or accountability that underlie academic integrity policies (Khlaif et al. 2025; Cotton, Cotton, and Shipway 2024; Wise et al. 2024; Khosrowi, Finn, and Clark 2023). Moreover, these rationalizations may overlook the fact that the work delegated to AI is often central to the process of learning to write. This process involves knowledge transformation by producing one’s own drafts and translating key concepts into sentences and paragraphs, thereby developing the ability to select appropriate vocabulary and to organize ideas effectively (Scardamalia and Bereiter 1987; Flower and Hayes 1981). Overall, students’ moral reasoning about AI use appears to be fluid, unfinished, and sometimes self-contradictory.

Our findings suggest that AI presents a steep, conceptual slippery slope for students that makes it difficult to uphold academic integrity. Because there are many available justifications to excuse every type and degree of AI use – from fixing minor grammatical errors, all the way to “vibe-writing” (Jelson et al. 2026) – extensive AI use is very easy to rationalize. Ethically relevant boundaries are hard to articulate and maintain, creating extreme ambiguity about what is morally acceptable behavior. (To be clear, we are not suggesting a *temporal* slippery slope, where minor AI use leads

over time to more extensive use, though that might also be happening. The slope is *conceptually* slippery (Govier 1982) – once one starts to rationalize, one finds rationalizations to fit any behavior.)

The slippery slope leads to a gap between faculty intention and student practice, and part of that phenomenon is explained by our five contexts of AI policy. Instructor intentions can be imperfectly expressed in formal policy; students can misinterpret a policy; they, of course, can also disagree with a policy; and, whatever their own policy is, they may not follow them. At each context, the boundary of acceptable AI use can be (intentionally or unintentionally) softened or relocated. The difference between adjacent contexts may be small (Govier 1982), but these differences can accumulate into a large gap between instructor intention and student practice. When the boundaries across contexts do not align, students can draw on the rationalizations in our taxonomy to treat each reinterpretation as insignificant.

We do not interpret students' rationalizations as evidence of moral failure. Indeed, our participants expressed considerable misgivings about the widespread use of AI among students, and often reported receiving unclear guidance about acceptable use or potential risks. The responsibilities are arguably as much with the corporations and universities that provide these powerful tools free to students, with little or no caution or guidance about their hazards, and to faculty and departments that have been slow to adjust. Meanwhile, students are caught in a rapidly changing pedagogical environment, expected to negotiate the boundaries of authorship, responsibility, and acceptable AI use with little else to go on other than brief, often out-of-date, instructions in course syllabi.

Our study findings require attention from faculty and educational institutions as they bring new insights into students' attitudes towards learning (including assignments), writing, and ethical practices in educational settings. The rationalization of **C6: Perceived Benefit** shows the role of learning goals in formal learning settings. Formal learning goals are often less visible to students because they are abstract and decontextualized (Brophy 2008). When the pedagogical rationale for an assignment is unclear, extensive AI use may appear to students not as a rejection of learning, but as a reasonable and efficient means of completing the task. Any formal learning environment arguably involves an implicit social contract in which the instructor helps the student learn, and the student is honest about what they know and do not know (Barnhardt and Ginns 2017; Gallant 2008). (It bears noting that this social contract is not between equal parties, because the consequences of contract violation are often far greater for students than for faculty.) The contract becomes difficult to sustain when learning goals and the reasons for restrictions on AI use remain unclear or irrational to students. Because AI blurs the boundary between student-written and AI-generated work, it becomes difficult for instructors to assess students' learning and provide meaningful feedback. A key intervention for higher education, then, is to help students understand explicit pedagogical rationales with course learning goals and AI policies, including why honest representation of one's own knowledge is important

for both meaningful learning and assessment.

The taxonomy presented in Tables 1 and 2 can be used to help students reflect on their own reasoning and help instructors explain why particular forms of AI use may conflict with course learning goals, thereby moving from unfinished rationalizations toward more deliberate rationales. Altogether, educational interventions and AI policies should be grounded in cognitively supportive approaches that develop more sustainable forms of AI use (and non-use) in academic contexts, while helping students critically reflect on their behavior (Adnin et al. 2025).

Limitations and Future Work

Our study has several limitations. First, our findings may not represent the AI-related moral perspectives of broader student populations or writing professionals whose work centers on writing, such as researchers and creative writers, as we interviewed only 20 undergraduate students who self-reported using AI for writing assignments. Future research could examine the roles of AI and ethical considerations across diverse writing domains and among a wider range of writing communities. Second, we did not interview instructors. As a result, we cannot determine how instructors intended students to interpret or apply their classroom AI policies. Future work could examine instructors' perceptions of AI policy, their pedagogical intentions, and their responses to students' rationalizations. Third, although we aimed to understand students' actual practices by collecting interview data alongside chat logs, course syllabi, and submitted assignments where available, not all participants provided these materials, limiting our ability to fully capture their psychological and behavioral processes of AI use. Future work could adopt different methods, such as screen-recorded writing sessions, to gather more comprehensive data.

Conclusion

Through semi-structured interviews with college students in the United States, we present new insights into students' rationalizations around the use of AI for academic writing. We identify five conceptual contexts where students negotiate what counts as appropriate AI use, and we show how divergence across these contexts contributes to a conceptual slippery slope. Our key findings suggest that students use 23 distinct rationalizations to explain why using AI in writing can be ethically acceptable. We also show that some of these rationalizations are used to justify AI use that violates course policies. Students also struggle internally with moral conflicts, and their rationalizations often emerge post hoc. These and other findings suggest that AI policy and educational practice need to do more than establish rules; they should also provide students with cognitive support for navigating ethical ambiguity.

Acknowledgments

The authors sincerely thank all student participants for their time and willingness to share their honest views on the use of AI in writing practices at the university.

References

- Adnin, R.; Pandkar, A.; Yao, B.; Wang, D.; and Das, M. 2025. Examining Student and Teacher Perspectives on Undisclosed Use of Generative AI in Academic Work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–17.
- Alexander, L.; and Moore, M. 2007. Deontological ethics.
- Ammari, T.; Chen, M.; Zaman, S.; and Garimella, K. 2025. How students (really) use ChatGPT: Uncovering experiences among undergraduate students. *arXiv preprint arXiv:2505.24126*.
- Ardito, C. G. 2025. Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2025(182): 11–28.
- Balalle, H.; and Pannilage, S. 2025. Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, 11: 101299.
- Bandura, A. 2011. Moral disengagement. *The encyclopedia of peace psychology*.
- Barnhardt, B.; and Ginns, P. 2017. Psychological teaching-learning contracts: Academic integrity and moral psychology. *Ethics & Behavior*, 27(4): 313–334.
- Bittle, K.; and El-Gayar, O. 2025. Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4): 296.
- Bodei, S.; Brumby, D.; Fisher, K.; and Mella, J. 2026. Co-Writing with AI: An Empirical Study of Diverse Academic Writing Workflows. In *Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work*, 1–16.
- Brophy, J. 2008. Developing students' appreciation for what is taught in school. *Educational psychologist*, 43(3): 132–141.
- Chan, C. K. Y. 2023. A comprehensive AI policy education framework for university teaching and learning. *International journal of educational technology in higher education*, 20(1): 38.
- Chan, C. K. Y. 2025. Students' perceptions of 'AI-giarism': Investigating changes in understandings of academic misconduct. *Education and Information Technologies*, 30(6): 8087–8108.
- Clarke, V.; and Braun, V. 2017. Thematic analysis. *The journal of positive psychology*, 12(3): 297–298.
- Cotton, D. R.; Cotton, P. A.; and Shipway, J. R. 2024. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, 61(2): 228–239.
- Ditto, P. H.; Pizarro, D. A.; and Tannenbaum, D. 2009. Motivated moral reasoning. *Psychology of learning and motivation*, 50: 307–338.
- Eaton, S. E. 2021. *Plagiarism in higher education*. Bloomsbury Publishing.
- Eaton, S. E. 2023. Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity*, 19(1): 23.
- Emig, J. 1977. Writing as a mode of learning. *College Composition & Communication*, 28(2): 122–128.
- Fleckenstein, J.; Meyer, J.; Jansen, T.; Keller, S. D.; Köller, O.; and Möller, J. 2024. Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Computers and Education: Artificial Intelligence*, 6: 100209.
- Flower, L.; and Hayes, J. R. 1981. A cognitive process theory of writing. *College Composition & Communication*, 32(4): 365–387.
- Foltynek, T.; Bjelobaba, S.; Glendinning, I.; Khan, Z. R.; Santos, R.; Pavletic, P.; and Kravjar, J. 2023. ENAI Recommendations on the ethical use of Artificial Intelligence in Education.
- Fu, Y.; Lin, Y.; Wang, Y.; Tran, S.; and Hiniker, A. 2026. "Everyone's using it, but no one is allowed to talk about it": College Students' Experiences Navigating the Higher Education Environment in a Generative AI World. *arXiv preprint arXiv:2602.17720*.
- Furze, L.; Perkins, M.; Roe, J.; and MacVaugh, J. 2024. The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4): 38–55.
- Gallant, T. B. 2008. Academic Integrity in the Twenty-First Century: A Teaching and Learning Imperative. ASHE Higher Education Report, Volume 33, Number 5. *ASHE higher education report*, 33(5): 1–143.
- Giray, L.; Jacob, J.; Encanto, V.; and Mansilungan, C. J. 2026. Cheating Writing with Generative AI: Exploring Student Motivations Using the Theory of Planned Behavior. *Journal of Academic Ethics*, 24(1): 1–23.
- Gonsalves, C. 2025. Addressing student non-compliance in AI use declarations: implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, 50(4): 592–606.
- Govier, T. 1982. What's Wrong with Slippery Slope Arguments? *Canadian Journal of Philosophy*, 12(2): 303–316.
- Gruenhagen, J. H.; Sinclair, P. M.; Carroll, J.-A.; Baker, P. R.; Wilson, A.; and Demant, D. 2024. The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence*, 7: 100273.
- Haidt, J. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, 108(4): 814.
- Hawdon, J.; Costello, M.; and Reichelmann, A. V. 2026. Cheating with ChatGPT and techniques of neutralization. *Deviant Behavior*, 47(4): 664–685.
- Huang, J.; and Chang, K. 2024. Citation: A key to building responsible and accountable large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*, 464–473.
- Hursthouse, R.; and Pettigrove, G. 2000. On virtue ethics.
- Jelson, A.; Manesh, D.; Jang, A.; Dunlap, D.; Kim, Y.-H.; and Lee, S. W. 2026. An empirical study to understand how

- students use ChatGPT for writing essays. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 1–26.
- Khlaif, Z. N.; Alkouk, W. A.; Salama, N.; and Abu Eideh, B. 2025. Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, 15(2): 174.
- Khosrowi, D.; Finn, F.; and Clark, E. 2023. Diffusing the creator: Attributing credit for generative AI outputs. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 890–900.
- Liang, W.; Yuksekgonul, M.; Mao, Y.; Wu, E.; and Zou, J. 2023. GPT detectors are biased against non-native English writers. *Patterns*, 4(7).
- Lin, X.; Huang, H.; Shi, S.; and Vines, J. 2026. Relying on LLMs: Student Practices and Instructor Norms are Changing in Computer Science Education. *arXiv preprint arXiv:2602.05506*.
- Maramark, S. 1993. *Academic dishonesty among college students*. US Department of Education, Office of Educational Research and Improvement, Office of Research.
- McCutchen, D. 1996. A capacity theory of writing: Working memory in composition. *Educational psychology review*, 8(3): 299–325.
- McCutchen, D. 2000. Knowledge, processing, and working memory: Implications for a theory of writing. *Educational psychologist*, 35(1): 13–23.
- Parker, L.; Loper, A. J.; Carter, C. W.; Hayes, J.; and Karakas, A. 2026. Longitudinal insights into AI in education: Usage, ethics, and policy development in higher education. *Computers and Education Open*, 10: 100329.
- Perkins, M. 2023. Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2): 1–24.
- Perkins, M.; Furze, L.; Roe, J.; and MacVaugh, J. 2024. The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6): 49–66.
- Perkins, M.; and Roe, J. 2024. Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy*, 37(3): 633–653.
- Scardamalia, M.; and Bereiter, C. 1987. Knowledge telling and knowledge transforming in written composition. *Advances in applied psycholinguistics*, 2: 142–175.
- Scarfe, P.; Watcham, K.; Clarke, A.; and Roesch, E. 2024. A real-world test of artificial intelligence infiltration of a university examinations system: A “Turing Test” case study. *PloS one*, 19(6): e0305354.
- Scheffler, S. 1988. *Consequentialism and its Critics*. Oxford University Press.
- Slimi, Z. 2026. A systematic critical review of generative AI’s impact on authorship, pedagogy, and integrity (2023–2025). In *Frontiers in Education*, volume 11, 1769680. Frontiers Media SA.
- Stearns, L. 1992. Copy wrong: Plagiarism, process, property, and the law. *Calif. L. Rev.*, 80: 513.
- Theoharakis, V.; Mylonopoulos, N.; and Papadopoulou, K. 2025. AI’s learning paradox: how business students’ engagement with AI amplifies moral disengagement-driven misconduct. *Studies in Higher Education*, 1–18.
- Tindle, R.; and Longstaff, M. G. 2015. Writing, reading, and listening differentially overload working memory performance across the serial position curve. *Advances in Cognitive Psychology*, 11(4): 147.
- Waqas, M.; Hania, A.; and Chunyan, X. 2026. Understanding AIgicism in higher education: the lens of general AI attitudes and moral disengagement. *Studies in Higher Education*, 51(4): 910–926.
- Wasi, A. T.; Islam, M. R.; and Islam, R. 2024. LLMs as writing assistants: Exploring perspectives on sense of ownership and reasoning. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants*, 38–42.
- Wise, B.; Emerson, L.; Van Luyn, A.; Dyson, B.; Bjork, C.; and Thomas, S. E. 2024. A scholarly dialogue: Writing scholarship, authorship, academic integrity and the challenges of AI. *Higher Education Research & Development*, 43(3): 578–590.
- Wu, C.; Zhang, H.; and Carroll, J. M. 2024. AI governance in higher education: Case studies of guidance at Big Ten universities. *Future Internet*, 16(10): 354.