Canadian Perspectives on Academic Integrity, *8*(4), 1-9.
http://doi.org/10.55016/ojs/cpai.v8i4/81748
Practitioner Article

Neutralizing the "Threat" of Technology: A Practical Guide for Re-Evaluating Assessments to Maintain Academic Integrity

Cory D. Scurr

Abstract
Technology, particularly generative artificial intelligence (GenAI), poses ostensible "threats" to higher education, especially to learning outcomes and academic integrity. This paper presents a "toolkit" that integrates four grounding concepts (cognitive offloading, automaticity, authentic assessment, and evaluative judgment). Educators can use the toolkit as a foundation to evaluate their current assessments and consider new assessments. Following this, the paper presents three "filters" or consideration frameworks to help neutralize GenAI's threat to assessments: *Abandon, Monitor, Enhance*, and *Adopt to Neutralize*. Each filter encourages educators to consider specific questions and contexts as they explore the value of their current assessments and determine how to best move forward in the wake of ubiquitous technology.

Keywords
academic integrity; assessment design; authentic assessment; automaticity; Canada; cognitive offloading; Evaluative Judgment; higher education; GenAI

Author affiliation: Conestoga College
Corresponding author: cscurr@conestogac.on.ca

Contextualizing the "Threat"

Technology has the power to disrupt learning. The higher education sector is often reactionary to ostensible "threats" that disrupt long-standing ways of doing or upholding academic integrity. Recently, generative artificial intelligence (GenAI) has been viewed as the main culprit of disruption as it has sent educators, staff, and administrators scrambling to ensure assessments are not compromised and remain valid measures of learning. The threat, then, is to both assessment security and intended learning outcomes. Finding GenAI's place within higher education is important. However, other factors may compromise assessments, including additional ways students support or bypass their learning. As a result, we need to have a bigger conversation, one in which GenAI is firmly situated within the context of sound teaching and learning and curriculum considerations.

Openo (2024) explains how the paradigm in the world of academic integrity has shifted away from a policy-driven, policing perspective to a view that assigns greater responsibility to the conditions of the teaching and learning environment. In other words, academic integrity must not be dominated by rule compliance, nor can the onus be placed squarely on students. Rather, the responsibility is shared with the instructional design of courses, faculty teaching practices, and program quality assurance processes (Bertram Gallant, 2008). In particular, attention must be paid to how assessment design influences student behaviour (Openo, 2024).

In the face of threats to assessment validity, one approach is to try to make expectations clear to students. Corbin et al. (2025) outline *discursive changes* to assessment, defined as "modifications that rely solely on communication of instructions, rules, or guidelines to students, such that their success depends entirely on student awareness" (p. 5). This approach is limited, however, as it relies on student compliance. Instead, Corbin et al. (2025) suggest *structural changes* to assessments or "modifications that directly alter the nature, format, or mechanics of how a task must be completed, such that the success of these changes is not reliant on the student's understanding, interpretation, or compliance with instructions" (pp. 6-7). *Structural changes* are not formulaic or prescriptive but do require a "conceptual toolkit by which instructors can understand what might count as appropriate assessments for their students" (Corbin et al., 2025, p. 9).

A Toolkit for Neutralizing the "Threat" to Learning and Academic Integrity

What follows is a toolkit, comprised of four grounding concepts and three "filters" or considerations, to determine how to best move forward (or not) with assessment revisions (see Figure 4 on p. 9). These frameworks essentially approach or review assessments from a "threat reduction" perspective. In *Defining Assessment Security in a Digital World*, Dawson (2021) suggests that the field of academic integrity is uncomfortable "with the systematic study of the ways students cheat," yet we must "both promote academic integrity and seek to detect cheating" (p. 3). I agree. Building on this idea, we must understand *how* and under *what conditions* students cheat to better navigate the process of designing valid assessments. In other words, we must dive into the muddy waters of cheating trends to reconceptualize how we measure students' learning so we can better and more confidently promote academic integrity and sound learning.

At first, this perspective may appear cynical or dejected; however, understanding how students can undermine or bypass assessment expectations and learning outcomes by cheating provides a solid starting point for course correction. Better yet, institutions should leverage academic misconduct data if they collect it to ascertain *how* students cheat, in which classes and/or programs higher rates of cheating occur, and at what point during the semester cheating happens most frequently. Of course, data that shows higher rates of cheating does not necessarily mean more cheating actually occurred compared to other courses; instead, it could simply mean more cheating was investigated and reported. That said, when used properly, academic misconduct data is a powerful tool for identifying cheating trends, which can (dare I say should) trigger assessment intervention or redesign. Thus, misconduct data may be useful to consider in conjunction with the concepts and filters presented below.

An Outcomes-Based Approach and Backward Design

Considering assessment design within the broader context of Program Learning Outcomes and Course Learning Outcomes is advantageous, since assessment design in isolation may lead to poor design or an assessment that expects too much or too little from students, depending on where they are in their program. Broadly speaking, it is wise to consider the program's terminal outcomes, then work backward. As Wiggins & McTighe (2005) explain,

> Our lessons, units, and courses should be logically inferred from the results sought, not derived from the methods, books, and activities with which we are most comfortable. Curriculum should lay out the most effective ways of achieving specific results . . . In short, the best designs derive backward from the learnings sought. (p. 14)

Programs themselves are organizational and educational, with predetermined structures and standards that, when completed, signify that the institution has certified the student as having completed the intended outcomes (Bens, 2024). As we navigate the nascent integration of AI into the curriculum, considering where AI fits into overall learning goals is crucial. This should go beyond learning outcomes that focus on the ethical and societal implications of AI and integrate *how* to ethically *use* it, particularly in the context of a specific subject area. Like how we teach students in a lab to use a pipette, so too should we teach students to wisely and effectively use AI in their field.

Coordinating assessment at the program level can promote and enable ethical assessments and academic integrity (Bens, 2024). Thus, those engaging in assessment (re)design should consider the program as a whole. Striving to create program cohesion through course and assessment design can effectively support student learning by having them "build on and reinforce one another" (Suskie, 2009, p. 4). Program considerations should not be isolated from industry and social needs, and institutions have an obligation to ascertain how GenAI is being leveraged and consider how AI literacy can be scaffolded into learning outcomes.

Grounding Concepts

Perspectives differ on whether GenAI compromises or enhances learning in individual courses or in programs. However, exploring the concepts of cognitive offloading, automaticity, authentic assessment, and Evaluative Judgment can support decision-making about if, how, and when technology, such as GenAI, can and should be used in teaching and learning contexts.

Concept 1: Cognitive Offloading

The concept of cognitive offloading is a good starting point to understand how various tools and technologies may compromise intended learning outcomes and when they are the right fit for the tasks and contexts at hand. Cognitive offloading is described as "the use of physical action to alter the information processing requirements of a task to reduce cognitive demand" (Risko & Gilbert, 2016, p. 676). Tools and technologies for cognitive offloading can range from pen and paper, language translation tools, calculators, auto-paraphrasing tools, formularies, and, more recently, GenAI. Dawson (2021) aptly explains that "cognitive offloading poses serious challenges for assessment, as educators and institutions are often unclear about which cognitive [offloading tools] are allowed and which are not" (p. 11). Oakley et al. (in press) caution that neuroscience-based research suggests shifting "away from explicit content instruction and memorization, combined with increased reliance on external memory aids [...] has actively contributed to declining cognitive performance" (p. 2). In other words, we must be thoughtful and considerate of what kinds of cognitive offloading tools we use and when to use them. Tools and technologies used to offload some of the cognitive effort required to complete complex tasks are not inherently threatening; in certain situations, some are excellent for enhancing learning and enabling higher-order cognition. Crucially, then, the intended assessment, course, and program learning outcomes must be identified to determine which cognitive offloading tool(s) may enhance or undermine the desired outcomes.

Concept 2: Automaticity

According to Bargh et al. (2012), the study of automaticity has permeated nearly all domains of psychological research, resulting in various definitions. *Learned automaticity* is created through the repeated execution of a behaviour in response to a trigger (Servan-Schreiber et al., 1996). Microsoft Copilot aptly summarized automaticity as ”the ability to perform tasks without conscious thought or effort, often as a result of extensive practice and repetition” (Microsoft, 2024). Like cognitive offloading, automaticity frees up cognitive resources, allowing us to focus on more complex tasks. This concept is important to consider when contemplating assessment design and desired outcomes. While we may relish the idea of our students recalling definitions, equations, or author names, doing so may not be necessary for the intended learning outcome. Other times, however, students absolutely must develop an automatic response. Thus, we must consider what we truly want our learners to demonstrate and whether that needs to come *automatically*.

Concept 3: Authentic Assessment

Authentic assessment can be regarded “as an umbrella term that seeks to immerse learners in environments where they can gain highly practical, lifelong learning skills” (Barkatsas & McLaughlin, 2021, p. ix). For instance, students can engage in project-based learning, where they create tangible products or presentations, or portfolio assessments that showcase their progress and achievements over time. Authentic assessments do not need to replicate *to the full extent* the knowledge and skills required in a real-world application. Assessment can be considered authentic even if only a *portion* of the relevant knowledge and skill set is being evaluated, as long as that portion will eventually be incorporated into a larger set of skills or knowledge required in the real world. Given the integration of AI, including GenAI, workforce requirements are not the same as they were 10 years ago or even more recently. As Openo (2024) states, “The internet and artificial intelligence ... have reduced the value of certain forms of assessment that do not teach truly human skills” (p. 223). As AI infuses evermore into our lives and the labour force, we must teach our students complex communication skills and how to navigate unstructured problems, like problem-finding and problem-framing (Reich, 2020; Openo, 2024).

Authentic assessments “create learning environments that neutralize cheating, to some extent, by engaging students in meaningful academic work grounded in real-world situations” (Openo, 2024, p. 219) and shift the focus from the cops-and-robbers mentality of academic misconduct to teaching and learning that promotes academic integrity (Openo, 2024). Authentic assessments, then, are crucial not only because they are more difficult to cheat on than some other forms of assessment but because they more accurately reflect the learning outcomes that are directly linked to our evolving labour market (Openo, 2024). Because authenticity has long been recognized as a feature of good assessment design (Bretag et al., 2020), asking oneself, “How authentic is my assessment?” is good and serves as a critical jumping-off point for revision. Authentic assessment, however, is not a panacea to academic misconduct, particularly contract cheating (Ellis et al., 2020), and must be considered in conjunction with other important concepts.

Concept 4: Evaluative Judgment

Evaluative judgment is the ability to make decisions about the quality of one’s and others’ work (Tai et al., 2018), including outputs from humans or machines. In the age of rapid technological advancements, we must evaluate what is good, what is bad, and why; as such, teaching students how to do this should be integrated into our learning outcomes. This relates to the necessity of authentic assessment, as our students and the labour force at large will require the ability to judge the quality of one’s own work, the work of other humans and machines, and appraise and understand standards of quality. Chen et al. (2022) explain that “as knowledge becomes readily available and quickly outdated, students need to be able to assess the quality of information provided to them and to evaluate their own responses and actions” (p. 493). We can help students develop evaluative judgment by providing them with more opportunities to see and evaluate their own work and their peers’ work. Activities that may promote these skills include self and peer assessment and engagement with exemplars, with discussion of grading criteria (Tai et al., 2018; Molloy et al., 2020). On top of this, we need to provide students with opportunities to engage with AI outputs in specific subject areas.

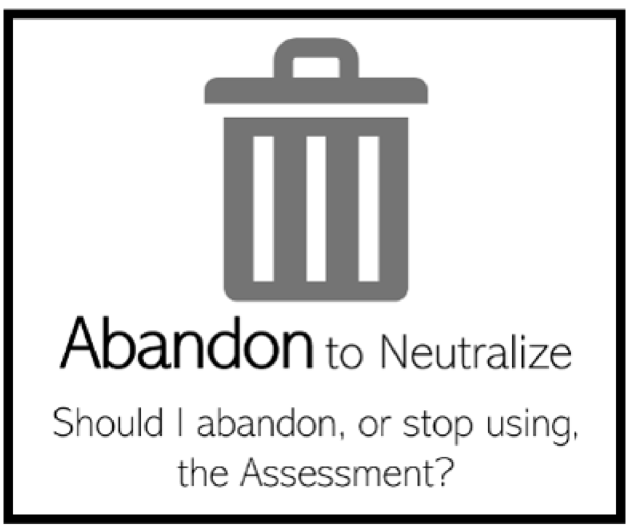
Three “Filters” to Support Assessment (Re)design

Deciding whether assessments require change can be difficult. What follows are three “filters,” or evaluation frameworks, to consider for assessment (re)design. These filters are intended to help neutralize “threats” to learning outcomes. Some assessments may need to be *abandoned*, whereas others may remain as is but require us to *monitor* students while they complete them. Other assessments will need to be *enhanced* through various means, such as *adopting* the tool or technology that may appear to compromise the intended learning outcome. Keeping the concepts of cognitive offloading, automaticity, authentic assessment, and evaluative judgment top of mind, these filters can support educators in determining what to do with assessments when faced with potential “threats” provoked by tools and technology when leveraged systematically.

Filter 1: Abandon to Neutralize

Shaw (2018) suggests that "the rise of the term paper happened to coincide with the shift in the approach to referencing in professional research writing," and students were ultimately expected to demonstrate their knowledge in a way that *mirrored that of the academy* (p. 127); otherwise the writing was deemed plagiarism. Most students will not become academics, so mirroring academic practices may not be necessary, beneficial, or authentic. To suggest "that is how we have always done it" or "that is what I did when I was in school" does not constitute sound logic for keeping an assessment. Understanding where assessments came from and why we use them can help us determine if they are still valid for measuring student learning. Applying the Abandon to Neutralize filter requires us to consider questions like when and why this assessment was created, what level of thinking we are striving for, and whether technologies should be leveraged to offload cognitive effort to make room for other tasks (see Figure 1). In addition, consider how industry or related professional areas use technologies. Potentially outdated assessments are those that

Figure 1. Abandon to Neutralize



require the demonstration of knowledge and skills that are no longer relevant in today's world because they have been replaced by technology in industry. For example, knowing how to use paper-based dictionaries may no longer be relevant as definitions can be found more quickly using search engines. Certain assessment types (e.g., memorizing definitions) are more subject to "threats" as information is nearly always at our fingertips (e.g., search engines), therefore, memorizing is not as important as knowing the importance or application of the information.

Various tools and technologies are pushing us to reconceptualize how humans communicate, or at least the *process* of this communication. While potentially inflammatory, we will likely need to rethink how we *expect* students to communicate. One of the six tenets of postplagiarism is that hybrid human-AI writing will become normal, and detecting where the human ends and the AI begins will be/is pointless and futile (Eaton, 2021). As a result, educators will need to adapt to what we expect to be reasonable and acceptable methods of communication, similar to how we most often *expect* students to solve complex mathematical problems using a calculator. To reiterate, perhaps we should be reconceptualizing what we *expect* from students regarding written communication.

Filter 2: Monitor to Neutralize

Broadly speaking, monitoring to neutralize refers to a faculty member watching or observing their students engage in an assessment to mitigate cheating (see Figure 2).

Figure 2. Monitor to Neutralize

At first glance, the "Monitor to Neutralize" filter may feel like a "dejected" lens or appear "Orwellian," but consider how learning often happens outside of a formal educational setting. Let's break down the knowledge, skills, and tasks of a stone mason teaching an apprentice how to dry lay a stone wall. The well-practiced mason will first demonstrate what to do, often providing verbal instructions. The apprentice will try to mimic the actions of her mentor, while the mason watches, or "monitors," the apprentice's progress. Monitoring skill demonstration is even more crucial when automaticity is necessary. Consider a pilot's ability to safely land a plane or navigate severe turbulence. This must come automatically because quick reaction time is essential. Student pilots must be monitored and tested (whether via a flight simulator or under the guidance of a flight instructor) to ensure they can automatically use their instruments correctly and execute the necessary physical maneuvers to safely fly a plane, all seemingly "without thought." All this is to say that it is sometimes okay and natural to monitor learning, though it may not be "natural" to monitor students in certain situations. Thus, to accomplish these tasks successfully, practice, repetition, and oversight of the implementation of knowledge are required to ensure mastery.

Automaticity is a useful concept to consider when leveraging this filter. If automaticity of a skill is particularly important, monitoring the demonstration of that skill is crucial, and offloading this to an alternative assessment type may be problematic. When scrutinizing an assessment, stop and think whether students must be able to automatically demonstrate this skill or recall this knowledge. If the answer is a definitive "yes," then monitoring students is crucial to ensure the outcome is met. If the answer is "maybe" or "no," then you may need to rethink how you are assessing them; suggestions found in the *Enhance to Neutralize* filter may help.

It may be tempting to think we are deterring our students from misconduct (or validating learning) by "making our expectations clear," as though clarity of instructions is an ostensible form of deterrence or monitoring. However, applying approaches that attempt to categorize assessments according to permitted levels of AI use, leveraging AI assessment scales, or adopting student declarative approaches (i.e., requiring students to declare or disclose their AI use) are *discursive* changes to assessments and have serious limitations that rely solely on students' adherence to "rules" (Corbin et al., 2025). Therefore, when deciding whether to use this filter, avoid the temptation of applying *discursive changes* in hopes of this being a method of monitoring. Overall, this filter intends to encourage educators to rethink *why* they are monitoring students and whether more *structural changes* (Corbin et al., 2025) should be considered.

Filter 3: Enhance and Adopt to Neutralize

We can consider the Enhance to Neutralize portion of the filter from two lenses: (1) practising exceptional teaching and learning strategies and (2) enhancing evaluation or rubric criteria (see Figure 3). The Adopt to Neutralize aspect of this filter refers to adopting the tool or technology that may pose a threat to the assessment. For instance, rather than prohibiting AI or trying to control or limit its use, the assessment embraces it.

Figure 3. Enhance and Adopt to Neutralize

Enhancing Teaching Practices

Stoesz (2024) explains that

> extensive research training in a particular field or content area is often the key qualifier for teaching in higher education, and faculty and contract (or sessional) instructors are often not required to complete formal training in andragogy or pedagogy (Crider, 2023; Murtonen & Vilppu, 2020). (p. 208)

While mandatory training is increasing in various subsectors of higher education, some faculty still may not have the necessary tools to feel competent and confident in their teaching, let alone their ability to (re)conceptualize assessments. Add to that the looming "threat" of GenAI to learning, and some faculty may feel hamstrung. Equipping faculty with tools and training is crucial to ensure that the paradigm shift (as described earlier) to a teaching and learning approach holds strong. While crucial, this is not the space to examine exceptional teaching practices and their influence on mitigating academic misconduct.

Enhancing Evaluation Criteria

For various reasons, a faculty member may not want or be able to change their assessment, so considering how we communicate our goal with a rubric is a better option. Not only can clear rubric criteria discourage AI use (if that is the goal), but it also helps to reframe the unauthorized use of AI as a lack of assessment validity rather than a moral and ethical consideration that requires an academic integrity investigation. In *Validity Matters More Than Cheating*, Dawson et al. (2024) suggest that

> from a validity perspective, when a student is found to have completed a task in a way that makes their assessor's judgment invalid, awarding the student no credit for the relevant parts of the task is simply assessment, not punishment. (p. 7)

In other words, if a student uses a tool or technology (e.g., AI) to complete an assessment when it was prohibited, then the assessment of the student's ability to demonstrate the learning outcome(s) is compromised. While the student will likely do poorly or fail that assessment, an academic integrity investigation is unnecessary, for which faculty will surely be pleased. To do this, sound evaluation criteria or rubrics are useful.

Paulson and Sharpe's (2025) "AI Savvy Rubrics for Writing Assignments" is a fine example of how we can action a validity perspective, whereby faculty are not changing their assessment but rather their evaluation criteria to emphasize the validity of learning outcomes where AI was prohibited. For instance, Paulson and Sharpe (2025) suggest we can discourage AI use by framing our metrics through poor standards and assignment context. Rubric criteria that consider poor standards might look like the following:

- *Information Accuracy*: inaccurate content will not meet the minimum standard
- *Verifiable References*: hallucinated or invented references will not meet the minimum standard
- *Language, Voice, and Style*: embellished, flowery, exaggerated language will not meet the minimum standard
- *Emphasis on Nuance or Complexity*: lack of addressing nuance or complexity will not meet the minimum standard

Rubric criteria that consider assignment contexts might look like the following:

- *Connections to Course Material*: vague, inaccurate, or general attempts to connect to specific course materials will not meet the minimum standard
- *Connections to Hyper-Local Issues or Organizations*: vague, inaccurate, or general attempts to relate to a specific issue, organization, current event, etc. will not meet the minimum standard
- *Evidence of Progression*: limited or incomplete evidence of drafts, notes, outlines, etc. do not meet the minimum standard
- *Class Activity or Discussion Synthesis*: a limited, inaccurate, or incomplete reference to class activities or discussion will not meet the minimum standard
- *Application of Human Feedback*: minimal application of specific peer or faculty feedback will not meet the minimum standard (Paulson & Sharpe, 2025)

Of course, faculty must massage or adjust their standards according to their courses and expected outcomes, but these rubric considerations are a solid jumping-off point for enhancing evaluation criteria.

Adopting Tools and Technologies

Educators may have polarized opinions about the value of adopting to neutralize. Adopting tools and technology simply for the sake of it may not be the best option. If we want to adopt AI or other technologies in our assessments, we must consider a few things. Firstly, we should ask ourselves, "Is industry using the tool or technology?" or "Do students need to know how to use the technology at the next level of their education?" and "Are we doing our students a disservice by *not* teaching them how to use the tool or the technology?" If the answers are "yes," then consider adoption into your teaching and assessment processes. Secondly, we should ask ourselves how we intend to incorporate the technology and whether we will measure how effectively students critique or evaluate the function or output or if we will measure the students' 'final product' with the use of the tool/technology and how effectively it was used.

Incorporating Technology

If we plan to incorporate technology, we need to consider whether using technology (e.g., GenAI) is a distinct learning outcome or if the focus will remain on measuring the overall production of the content of the assessment (e.g., will the use of GenAI be considered holistically as part of the submission, similar to the incorporation of good sources). Ultimately, educators must decide on the end game of how they want their students to leverage the tool, then choose a trajectory.

We might need to (or dare I say we must) reconsider how we do history, or marketing, or psychology, or anthropology, etc. How we research, validate, and produce content will need to be reconsidered. It is important that students understand the difference between using AI as a research tool versus a content creation tool because the former requires different skills and parameters than the latter. Highlighting this difference and integrating it into curriculum considerations is an important aspect of teaching AI literacy.

Currently, it appears these two elements are combined by students who use AI. It is almost like cutting and pasting from the internet but on steroids. For instance, if a student were asked to write about the role of women in World War 1 in Canada, they could type the prompt into a search engine, then copy and paste various answers from sites like Veterans Affairs Canada or The Canadian Encyclopedia. In other words, the 'research' (in this instance, the rudimentary search for information) and the locating, cutting, and pasting of the information are separate actions. Now consider this same question, slightly adjusted to "*write me an essay* on the role of women in World War 1 in Canada," prompted into AI. The student will have 'conducted the research' and 'created the content' (in this case, an essay) immediately and effortlessly. The ability to mash together research and output is now virtually effortless, yet good quality is not guaranteed. In the age of AI, scaffolding and *intentionally separating* the research portion from the output portion of an assessment, then, is crucial.

Educating students on AI's ability to 'do research' or be used as a research tool is important and can be considered separately from prompt engineering and evaluating AI outputs. It may be useful to highlight AI's abilities like educators have highlighted Wikis: "It's an okay place to start, but ... " The question is, how do we turn this into a measurable learning outcome, or is it even necessary to treat it as a separate outcome? The struggle educators have now with students using search engines like Google, rather than their institution's library database, will only be exacerbated with AI. A key difference between these two is that AI provides a much more compelling illusion of truth. While a student will miss out on the depth of resources available by using Google and may end up with 'non-scholarly' or non-peer-reviewed sources, they do not end up with resources that simply do not exist at all, as can be the case with AI. We may want to consider these angles when deciding how to incorporate AI into our assessments from a research perspective.

Measuring Evaluative Judgment

Ideally, if educators adopt AI tools, then a scaffolded approach that focuses on the *process* of learning over the *product* is ideal. This includes teaching students how to use the tool successfully within the subject area. To do so, this means adopting a scaffolded approach to both the incorporation of the tool and the completion of the assessment itself. Adopting AI in this way is a good first step in introducing students to the capabilities of the tool and highlights its strengths and limitations relative to certain subject areas or tasks. In these situations, students use the tool less to create content that will be submitted for evaluation but rather create critiques or judgments on the outputs AI provides. As educators, we arguably have an obligation to show students what the tool is exceptional at doing, where it is lacking, and the grey areas in between, relative to our subject matter areas.

When AI is leveraged in these situations, it is the student’s *judgment of the outputs* that is being measured, rather than how students interact and use the outputs themselves. As Tai et al. (2018) explain, “As knowledge becomes readily available and quickly outdated, students need to be able to assess the quality of information provided to them and to evaluate their own responses and actions” (p. 493). This, I argue, can be applied directly to assessing AI outputs. In these assessment situations, the AI tool can be used to create content so our students can learn to effectively judge the AI tool’s output, and the content the AI tool provides should be subject-specific.

Students can engage with the output in a high-level way if they do not have background knowledge of the topic or subject, but the critique, or critical analysis, will inevitably be shallow. Without background knowledge or something to which students can juxtapose an AI output, we can ask students questions like, “What is your impression of the depth of knowledge the tool provided?”, “Did you like the way the content was organized?”, “Do you trust the output?”, or “Do you trust the sources the output provided?” This may serve as a worthwhile low-stakes or no-stakes exercise for students to begin answering questions as a means to strengthen their AI literacy, but this exercise will serve more as an opinion piece, rather than an in-depth critical analysis.

If we want to promote more advanced critical thinking, students must have engaged with content in some other way or have background knowledge of the output (besides learning about it through AI). Having other content or information to which the AI output can be juxtaposed is necessary. For instance, prompting AI to “provide a 250-word overview about whether it is ethical to prioritize the education of gifted students over those with learning difficulties or disabilities,” and asking students to judge the output is futile and pointless if they have no prior knowledge of the topic. However, if students first learn about, say, utilitarianism, distributive justice, or deontological ethics (either through in-class teachings, a podcast, or reading), and *then* ask them to critique the output from one of these perspectives, this would be fair.[1] We can then assess a student’s ability to exercise evaluative judgment. In this case, we would be assessing the student’s ability to critically analyze, not their use of the tool itself. This is an important step in a student’s ability to use the tool effectively, since, as we know, humans must be the final decision-makers about the quality of robot outputs.

Another assessment strategy that falls under this category is teaching a student to *assess prompts*. Co-pilot defines prompt engineering as

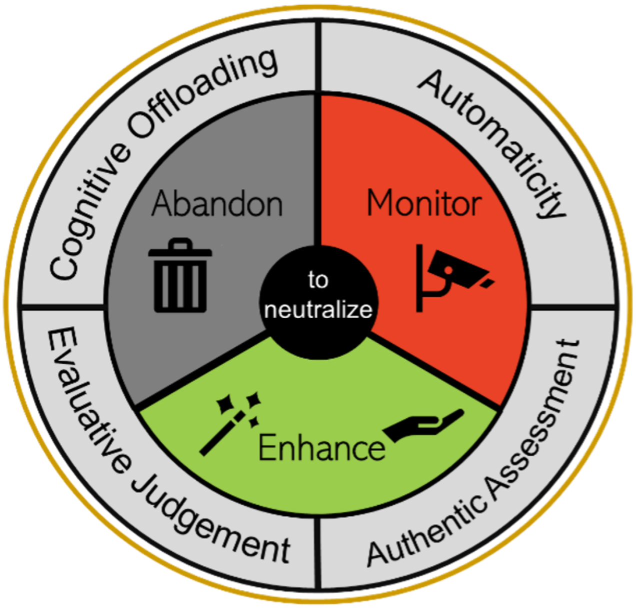
> the process of designing and refining prompts to guide large language models [. . .] in generating accurate and relevant outputs. This involves creating specific inputs that provide context and direction, ensuring the AI understands the nuances and intent behind the query. (Microsoft, 2024)

Chen et al. (2022) show that students can make complex judgments about question quality. If that is the case, I suggest we expand our expectations to *prompt quality* by asking them to evaluate the quality of their own and others’ prompts. This may serve as a good step forward in scaffolding as it still leverages evaluative judgment, but the student is beginning to learn how to use the tool to produce accurate outputs by judging what first goes in. Having students generate simple, one- or two-sentence prompts and compare those to multilayered prompts, on top of having them assess the various prompts relative to the outputs, is a good technique for building critical thinking through evaluative judgment while teaching AI literacy skills.

The onus, then, will be on faculty to think critically about how they want to incorporate AI into assessment and what it is specifically they want to measure. Resources and guidance have been developed for faculty to provide to students about their expectations of AI use, notably the AI Assessment Scale (AIAS) (Perkins et al., 2024). However, this is considered a discursive change and relies on student compliance (Corbin et. al, 2025). The above, as mentioned throughout the paper, is intended to provide faculty with practical considerations for how they want to incorporate AI. Simply stating students can or cannot use AI is not enough; we need to be more targeted with our assessment creation and learning outcomes. In other words, we must understand how students are using the tool and think about how we might compartmentalize AI use into separate, distinct learning tasks.

[1]To get this example, I asked Co-pilot the following: “What are some interesting prompts a student can ask about ethics?” One of the examples it provided was “Is it ethical to prioritize the education of gifted students over those with learning difficulties or disabilities?” I then asked, “What are key theories we would need to know if we wanted to answer the following question: ‘Is it ethical to prioritize the education of gifted students over those with learning difficulties or disabilities?’” It gave me a list, from which I chose the above three.

Figure 4. Toolkit for neutralizing "threats" to learning and academic integrity



Note: This figure ties together the grounding concepts of cognitive offloading, automaticity, authentic assessment, and evaluative judgment, with the Abandon, Monitor, and Enhance & Adopt filters, into a Toolkit for Neutralizing the "Threat" to Learning and Academic Integrity. Together, it is hoped that this will provide a useful framework for educators.

Conclusion

Re-evaluating student assessments in the age of rapidly evolving tools and technologies is not merely a reactive necessity but rather a proactive opportunity. As educators, our role is to foster environments where academic integrity is not enforced but inspired. Academic integrity is best preserved not through rigid control but through intentional alignment of assessment practices with authentic learning outcomes. Considering concepts like cognitive offloading, automaticity, authentic assessment, and evaluative judgment helps ground us to scrutinize what it is we are actually assessing, whether tools and technology should be introduced, and how. The three filters, Abandon, Monitor, and Enhance (& Adopt), are simple yet useful consideration frameworks that can guide faculty and faculty teams as they consider assessment (re)design in the age of AI. Let us embrace disruption as a catalyst for innovation, and let our assessments reflect the dynamic, complex, and human-centred nature of learning itself.

References

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, *16*(12), 593-605. https://doi.org/10.1016/j.tics.2012.10.002

Barkatsas, T., & McLaughlin, P. (2021). Preface. In T. Barkatsas & P. McLaughlin (Eds.), *Authentic assessment and evaluation approaches and practices in a digital era: A kaleidoscope of perspectives* (p. ix). Brill. https://doi.org/10.1163/9789004501577

Bens, S. (2024). Coordinating ethical assessment at the program level. In S. E. Eaton (Ed.), *Second handbook of academic integrity* (p. 233–248). Springer International Handbooks of Education. https://doi.org/10.1007/978-3-031-54144-5_91

Bertram Gallant, T. (2008). Academic integrity in the twenty-first century: A teaching and learning imperative. *ASHE Higher Education Report*, *33*(5), 1–143. https://doi.org/10.1002/aehe.3305

Bretag, T., Harper, R., Rundle, K., Newton, P. M., Ellis, C., Saddiqui, S., & van Haeringen, K. (2020). Contract cheating in Australian higher education: A comparison of non-university higher education providers and universities. *Assessment & Evaluation in Higher Education*, *45*(1), 25–139. https://doi.org/10.1080/02602938.2019.1614146

Chen, L., Howitt, S., Higgins, D., & Murray, S. (2022). Students' use of evaluative judgement in an online peer learning community. *Assessment & Evaluation in Higher Education*, *47*(4), 493–506. https://doi.org/10.1080/02602938.2021.1933378

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2025.2503964

Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. https://doi.org/10.4324/9780429324178

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, *49*(7), 1005-1016. https://doi.org/10.1080/02602938.2024.2386662

Eaton, S. E. (2021). *Plagiarism in higher education: Tackling tough topics in academic integrity*. ABC-CLIO.

Ellis, C., van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., … Saddiqui, S. (2020). Does authentic assessment assure academic integrity? evidence from contract cheating data. *Higher Education Research & Development*, *39*(3), 454–469. https://doi.org/10.1080/07294360.2019.1680956

Microsoft. (2024, September 20). *Copilot (gpt-4) [large language model]*. https://copilot.microsoft.com/

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assess-*

ment & Evaluation in Higher Education, *45*(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Oakley, B., Johnston, M., Chen, K.-Z., Jung, E., & Sejnowski, T. (in press). The memory paradox: Why our brains need knowlegde in an age of AI. In *The artifical intelligance revolution: Challenges and opportunities.* Springer Nature.

Openo, J. (2024). Academic integrity and the affordances and limitations of authentic assessment. In S. E. Eaton (Ed.), *Second handbook of academic integrity* (p. 217-231). Springer International Handbooks of Education. https://doi.org/10.1007/978-3-031-54144-5_90

Paulson, E., & Sharpe, A. (2025). *AI savvy rubrics for writing assignments.* Faculty Learning Hub, Conestoga College. https://tlconestoga.ca/ai-savvy-rubrics-for-writing-assignments/

Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context: A test of a theoretical model. *Archives of General Psychiatry*, *53*(12), 1105-1112. https://doi.org/10.1001/archpsyc.1996.01830120037008

Shaw, P. (2018). Are we making our students plagiarize? In D. Pecorari & P. Shaw (Eds.), *Student plagiarism in higher education: Reflections on teaching practice* (p. 125-135). Routledge. https://doi.org/10.4324/9781315166148

Stoesz, B. M. (2024). Academic integrity through ethical teaching and assessment: Overview and current trends. In S. E. Eaton (Ed.), *Second handbook of academic integrity* (p. 203-216). Springer International Handbooks of Education. https://doi.org/10.1007/978-3-031-54144-5_89

Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). Jossey-Bass.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*, 467–481. https://doi.org/10.1007/s10734-017-0220-3

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development.