

ALGORITHMIC EPISTEMIC OUTSOURCING AND EPISTEMIC CORRUPTION

Nimrod J. Dacles
Ateneo de Manila University, Philippines

The use of generative AI tools (also known as algorithmic epistemic outsourcing), such as ChatGPT, has become increasingly prevalent among college students. Much of this appeal stems from their promise of productivity and convenience, particularly in reducing the cognitive effort required for tasks such as essay writing. While these technologies can enhance efficiency, their unreflective use, characterized by a passive cognitive surrender, raises serious concerns about genuine learning, the development of relevant intellectual skills, and the cultivation of proper dispositions toward inquiry. More troublingly, habitual offloading of cognitive tasks to algorithmic artifacts runs the risk of fostering defective and undesirable characterological states, especially overreliance. By employing Ian James Kidd's language of epistemic corruption and his method called corruptionist criticism, I argue that the increasing practice of algorithmic epistemic outsourcing constitutes what might be called an epistemically corrupting condition that facilitates the acquisition of the epistemic vice of overreliance. I also argue that this form of epistemic corruption is not intrinsic. Rather, it emerges conditionally as a result of specific psychological and institutional factors. Finally, I explore possible ways to address the conditional nature of such a corrupting process by considering a pedagogical strategy (educating for intellectual virtues), ameliorative strategy (educating against intellectual vices), and a systemic strategy that involves adopting a more humanistic form of education that puts a premium on the individual student's growth as a thinker, learner, and human being.

Keywords: Epistemic corruption, overreliance, epistemic vice, algorithmic epistemic outsourcing, conditional, pedagogical strategy, ameliorative strategy, systemic change

INTRODUCTION

With the availability of generative artificial intelligence technologies like ChatGPT, concerns in relation to their usage have also emerged, like the one Radu Uszkai (2024) describes as “algorithmic epistemic outsourcing” (21), which refers to how students offload academic tasks such as researching, analyzing, or composing written output to algorithmic systems like ChatGPT. At first glance, practicing this seems advantageous for the student

since doing so enhances productivity and affords convenience, particularly because it reduces the time and cognitive effort required for completing academic tasks that traditionally required more time and were performed painstakingly through their own intellectual effort. However, when done unreflectively and passively to the point where they relinquish their own thinking and creativity; it then raises serious concerns about genuine learning and the development of relevant intellectual skills (Zhai et al. 2024; Sok and Heng 2023; Sternberg 2024; Krook 2025).

The problem here lies in the fact that in the learning process, students are expected to actively participate in activities such as reading, writing, reasoning, and reflecting, all of which require sustained attention, effort, and intellectual engagement. These activities are vital in that they provide opportunities for learners to develop their cognitive skills. To paraphrase a point raised by Flynn (2012, 2016), it is by confronting challenging cognitive activities that human skill is developed. However, when students use generative AI tools such as ChatGPT to outsource writing and related cognitive tasks, this epistemically formative process is compromised. And as a result, they get deprived of the essential activities that could have honed their skills. Scholars like Hadeel Naeem (2024) have described this phenomenon as deskilling, precisely because AI is depriving its users of the opportunity to exercise their essential human skills.

However, there is another troublesome consequence other than deskilling. The habitual offloading of cognitive task not only deprives its user from opportunities conducive for training essential skills but also runs the risk of making its user more susceptible to fostering epistemic vices—which are “blameworthy or otherwise reprehensible character trait, attitude, or way of thinking that systematically obstructs the gaining, keeping, or sharing of knowledge” (Cassam 2019, 23)—such as laziness, lack of oversight and most especially overreliance.

To analyze this vice and how it is facilitated by the unreflective practice of algorithmic epistemic outsourcing, this study will employ the language of epistemic corruption and the method called corruptionist criticism, both of which are coined and explored by the contemporary analytic epistemologist Ian James Kidd. Taking both as my conceptual and structural framework, I argue that the increasing practice of algorithmic epistemic outsourcing constitutes what might be called an epistemically corrupting condition that facilitates the acquisition of the epistemic vice of overreliance.

I begin with an exposition of Kidd’s concept of epistemic corruption in the section on Epistemic Corruption. After this expository presentation, the Corruptionist Criticism is carried out in the next section. First, I identify the corruptor and corruptee; second, I undertake the task of determining the epistemic vice promoted in the unreflective practice of algorithmic epistemic outsourcing. After determining the relevant subjects and objects of the criticism, I then focus on explaining how the unreflective practice of algorithmic epistemic outsourcing operates as a corrupting process and on assessing the strength of its corrupting tendency. To do so, I begin by describing the structure of writing as a learning activity, and then show how this structure is disrupted when writing is outsourced to generative AI systems like ChatGPT. I then turn to the psychological profile of students and some social-contextual factors to gauge the strength and determine whether the epistemically

corrupting condition is conditionally or intrinsically corrupting. Having a firmer grasp of the corrupting process, I then explore some possible corrective measures to the epistemic crisis posed by the unreflective practice of algorithmic epistemic outsourcing in the section on Conditionality and Corrective Claim.

WHAT IS EPISTEMIC CORRUPTION?

“Epistemic corruption occurs when one’s epistemic character comes to be damaged due to one’s interaction with persons, conditions, processes, doctrines, or structures that facilitate or provide conditions for the development and exercise of epistemic vices” (Kidd 2021, 71). This definition highlights that epistemic corruption can be triggered by a variety of factors, individually or in combination. It may occur through prolonged exposure to persons exhibiting vicious dispositions, oppressive social conditions that foster internalization of defective character traits, or both simultaneously. For example, prolonged subjection to gendered or racialized epistemic injustices can inhibit the development of intellectual virtues and character (Fricker 2007; Medina 2013).

Persons, conditions, processes, doctrines, or structures facilitate epistemic corruption by either providing an architecture or environment that deteriorates the pre-existing virtues already present in the subject’s character that in turn facilitates the development and exercise of epistemic vices or by providing an environment that starves the intellectual virtues thereby preventing them from being cultivated by those subjected to such conditions. Kidd determines both as two different kinds of damage which correspond to two types of corruption, namely: active corruption and passive corruption (Kidd 2021, 71). The first type is active as it proactively erodes the pre-existing or still developing virtues in the subject’s character while the second is denoted as passive however this shouldn’t be taken as a way of trivializing the damage incurred, the term passive implies that epistemically corrupting conditions can make it difficult for a person or agent exposed within it to develop his or her propensity for virtuous disposition, which consequently results to failure. The corruption is passive because the damage incurred does not involve the erosion of the qualities of the epistemic agent; nevertheless, he or she is hindered from developing epistemically virtuous qualities, which is detrimental to his epistemic character.

Kidd further identifies five modes of epistemic corruption: acquisition, activation, propagation, stabilization, and intensification, which describe various ways in which individuals become epistemically vicious as a result of their interaction with corrupting conditions (Kidd 2021, 72-73). In his vocabulary, activation occurs when a latent vice in a subject’s epistemic character is triggered. Acquisition, on the other hand, refers to the cultivation of novel epistemic vices that were previously absent from the subject’s epistemic character. The remaining three modalities of epistemic corruption differ from the two above in that they involve changes in the epistemic vices already present and active in the subject’s character (Kidd 2021, 72). With this in mind, let’s consider the other three modalities. Propagation refers to the way in which an epistemically corrupting condition widens the scope of a vice, from a localized form of close-mindedness to an expansive form of close-mindedness. Stabilization refers to

the modality of epistemic corruption that establishes fluctuating vices, rendering them more stable than they initially were and thus making them completely entrenched in the subject's intellectual character. Lastly, intensification occurs when existing vices become more severe in their manifestation. These five modalities don't necessarily operate in isolation, rather they can operate collectively (Kidd 2021, 73)

Aside from the distinction between active and passive and the different modalities of epistemic corruption, Kidd also clarifies some of its essential features and conditions. First, according to Kidd, epistemic corruption has no intentionality condition. This means that a system, process, or doctrine can produce corrupting effects even if corruption was never part of its design or purpose (Kidd 2019, 224). For example, a form of educational system can facilitate the acquisition of epistemic vices like close-mindedness to its constituents even if by virtue, it wasn't intentionally designed to do so. In this example, the corruption incurred is unintentional. Hence, a line of demarcation must be drawn to separate systems, processes or doctrines that deliberately undermine the intellectual character of their victims and those that do so unintentionally. Another important feature of epistemic corruption is that it has no success condition (Kidd 2019, 225). In other words, a system, process, or doctrine can be corrupting even when it fails to produce corruption in some of its members; as long as it has the tendency to undermine the intellectual character of anyone who gets exposed to it, then it is a corrupting condition (may it be a system, process or doctrine). In the case where some corruptees are untouched or unaffected by the corrupting tendency, it is merely incidental that their intellectual character and integrity are still intact. However, this does not remove the threat associated with such a corrupting condition. Kidd further clarifies that the strength of corrupting tendencies will depend on at least two factors: (1) the psychosocial profile of the potential victims and (2) the structural and normative features of the system (Kidd 2019, 225). Finally, an essential feature of epistemic corruption is that it operates in a dynamic way, combining a wide array of personal, contextual, and structural factors that facilitate, intensify, or entrench one or more epistemic vices (Kidd 2019, 225).

With such desiderata, my criticism will then have to account for whether the corrupting process entails active or passive epistemic corruption, whether the modality involves vice acquisition or vice intensification, or rather a combination of two or more modalities, whether the corrupting tendency of the corruptor is intentional or unintentional, and lastly, whether there are multiple factors that are involved in the corrupting process.

Having taken note of such considerations, we are now then in a better position to undertake the inquiry about how the practice of algorithmic epistemic outsourcing constitutes what might be called an epistemically corrupting condition by employing the method of corruptionist criticism.

CORRUPTIONIST CRITICISM

According to Kidd, an effective corruptionist criticism must explain and specify the following: (1) the corruptor(s) and the corruptee(s), (2) epistemic vice, (3) the

specificity and strength of the corrupting condition(s), and (4) the relevant conditionality and corrective claims (Kidd 2019, 227).

Hence, for this section my aim is fourfold: (1) to identify who or what is doing the corrupting and who or what is subjected to corruption, (2) pinpoint what specific vice is being inculcated or promoted in the process (3) lay down the specific processes and practices that facilitate epistemic corruption and assess the strength of the corrupting condition and (4) determine whether algorithmic epistemic outsourcing-induced epistemic corruption is conditionally or intrinsically corrupting (conditionality claim) and explore possible ameliorative projects to reduce the corrosive effect of such corrupting condition (corrective claims).

It must be noted that this fourfold method of corruptionist criticism employed in this paper was originally designed and used by Kidd to critique educational systems that facilitate corruption rather than edification¹. While such is the case, this, however, does not prevent the possibility of extending its applicability to other forms of corrupting conditions. Rather, I think of Kidd's framework as a general template to criticize system, condition, process, or any sort of corrupting condition. For this reason, I see no problem with applying his framework to describe, analyze, and evaluate how the unreflective practice of algorithmic epistemic outsourcing facilitates epistemic corruption.

Corruptor and Corruptee

The first task of a corruptionist criticism is to identify the corruptor and the corruptee, that is, to specify what is doing the corrupting and who or what is being subjected to epistemic corruption (Kidd 2019, 227). In the present case, the primary corruptees are students who regularly engage in algorithmic epistemic outsourcing in academic contexts, particularly in cognitively formative tasks such as writing. While it is possible that other agents, such as professionals or members of the general populace, may also be affected by similar practices, this paper restricts its analysis to students, who are plausibly more vulnerable to epistemic corruption due to their developing epistemic character (Kidd 2019, 225).

Identifying the corruptor requires a broader conception than one limited to systems as Kidd (2019) qualifies abstract entities such as processes as corrupting conditions as well. While he did not specify what he meant by process and provide an example of a corrupting process that I could use as a template, however, it seems accurate to interpret it to mean that epistemic corruption can also be facilitated by routinary, normalized, and seemingly harmless practices. With this in mind, the corruptor in this case is not generative AI technology as such, nor any particular AI system or developer, but the practice of algorithmic epistemic outsourcing which is understood as the passive delegation of epistemically formative cognitive activities like writing to generative AI systems. Here, my use of the term is closely similar to Shaw and Nave's (2026) account of cognitive surrender, which they describe as the "abdication of critical evaluation in which a user relinquishes cognitive control and adopts the AI's judgment as their own (4)."

What Specific Vice is Being Inculcated or Promoted in this Process?

Having identified the corruptor and the corruptee, we may now turn to the specific epistemic vice that is being inculcated or promoted through the practice of algorithmic epistemic outsourcing. In order to do so, it might be helpful to consider what the nature and structure of an epistemic vice is.

In the literature of vice epistemology, there are two competing proposals in conceptualizing the structure of an intellectual vice. On the one hand, Quassim Cassam's vice-obstructivist account, which conceives epistemic vices as reprehensible character traits, attitudes, or ways of thinking that systematically obstruct the acquisition, transmission, and retention of knowledge (Cassam 2019, 23). On this view, a trait qualifies as an epistemic vice insofar as it reliably interferes with epistemic goods, regardless of the agent's motivation. In this regard, one might say that vice-obstructivism bears a structural resemblance to virtue reliabilism², although the scope of its included vices in its repertoire is more expanded compared to the scope of virtues included in virtue reliabilism. But one plausible analogy of the two is how both does not require motivation. On the other hand, Jason Baehr's vice-motivationalist account which emphasizes motivation as constitutive feature of epistemic vice. Analogous to virtue responsibilism.³ This account holds that epistemic vices involve defective motivations, such as the lack of care toward epistemic goods, that dispose agents toward epistemically blameworthy conduct (Baehr 2019, 29). While it is not my interest to participate in this debate, this paper closely leans to obstructivism in understanding what constitutes an epistemic vice. In my defense, I say that this conceptual choice is motivated by the explanatory demands of the present inquiry.

With this background, I argue that the epistemic vice most saliently promoted by algorithmic epistemic outsourcing is overreliance, which "occurs when users excessively trust and depend on the model, potentially leading to unnoticed mistakes, inadequate oversight, loss of opportunity to develop new skills, and even the loss of important skills" (OpenAI et al. 2024, 59). The quoted text invites few comments. First, overreliance is a cancerous trait. It facilitates the emergence of further epistemic defects, such as carelessness and intellectual laziness. In other words, it can function as a gateway for other epistemic vices. Second, overreliance involves a passive engagement with an algorithmic artifact, which reflects what Duncan Pritchard (2019) has described as a non-extended manner of engaging with technology. And finally, overreliance involves excessive trust. This feature is particularly important in that how an agent relates to a source of information shapes how that source is integrated into one's epistemic practices. A high level of trust towards an informational source, for example, might lead an agent to overestimate the credibility and reliability of such a source, which in turn will result in a bad epistemic practice if the source provides faulty information. The study of Shaw and Nave (2026) substantiates this point. One of the key findings of their study is the observation that individuals tend to give a moderately high level of credence to a generative AI system (an informational source), given its confident and compelling outputs, which nonetheless can be false. Individuals would treat such an algorithmic artifact as epistemically authoritative, "lowering the threshold for scrutiny" (Shaw and Nave 2026, 45), and as a result, might lead to a bad epistemic practice.

Hence, I argue that overreliance as an epistemic vice involves a pattern of passive engagement with an algorithmic artifact due to excessive trust in the latter, which leads to mistakes, inadequate oversight, and loss of important cognitive skills, which in turn interfere with the acquisition, transmission, and retention of knowledge or other epistemic goods. How is this so? Inadequate oversight, for example, can result in an uncritical integration of faulty information into one's economy of beliefs, which in one way or another prevents one from acquiring knowledge. We can better understand this point by considering, for example, the case of David.

David is working on his term paper for his Buddhist Thought course. His chosen topic explores how Buddhist philosophical literature might shed light on the ethico-epistemological issue of fake news and misinformation. Having completed a preliminary draft containing the context, aim, and questions, he now turns to the task of engaging the relevant literature. To expedite the process, David decided to use ChatGPT to compile books and articles that have a direct bearing on his research question. Within seconds, the system automatically produces an extensive bibliography with a corresponding summary which appear compelling. Unbeknownst to David, though authoritative and convincing, some of the cited works are products of AI hallucination rather than actual publications. Yet because David has come to perceive ChatGPT as a trustworthy source, he does not see the need to verify the references against actual databases. Trusting the system's output, he integrates the bibliography into his draft as is.

Based on the case presented above, it becomes clear that when an agent fails to exercise a sufficient amount of oversight by cross-checking sources, the probability of incorporating falsehood into one's economy of beliefs increases, which in turn, prevents him or her from acquiring knowledge. Furthermore, the effect here is not isolated to an epistemic agent because once a faulty belief is integrated into one's web of beliefs, it can distort subsequent reasoning, inform further judgments, and be transmitted to others, in which case, the effect goes beyond the individual agent.

This section has established that the practice of algorithmic epistemic outsourcing facilitates the epistemic vice of overreliance. Or at the very least, it has demonstrated that the former is closely associated with the latter. The next section provides a more nuanced treatment of this topic by homing in on the actual process of epistemic corruption implied in the practice of algorithmic epistemic outsourcing.

How does GenAI Facilitate Epistemic Corruption?

This section describes how algorithmic epistemic outsourcing operates as a corrupting process and assesses the strength of its corrupting tendency. To do so, I begin by describing the structure of writing as a learning activity, and then show how this structure is disrupted when writing is outsourced to generative AI systems like ChatGPT. I then assess the strength of this corrupting condition by looking at the psychological profile of students and the normative structures of contemporary educational systems.

Ordinarily, when students engage in cognitively demanding activities like writing an essay, they need to exercise a range of higher-order thinking skills to accomplish this. Without a particular order in mind, writing involves creating an outline, developing and refining an argument, consolidating all the relevant information from one's knowledge, insight and experience, and organizing these materials into a coherent topic. Furthermore, it involves imagining alternative possibilities, considering objections, checking for consistency among the paragraphs, and spotting and revising for clarity and grammatical soundness. These process takes time and effort and sometimes can be painstakingly tedious but such process is necessary for the students to develop their independent thinking, creativity and other writing-related skills like copyreading. Through repeated engagement, students cultivate not only technical writing skills but also intellectual habits such as thoroughness, carefulness and autonomy. For this reason, writing plays an important role in the intellectual development of learners.

However, when students passively use generative AI tools such as ChatGPT to outsource writing and related cognitive tasks, this epistemically formative process is compromised. This displacement has a corrupting effect insofar as it deprives students the opportunity to practice the very activities through which intellectual virtues could have been cultivated. Over time, the practice of outsourcing weakens the student's engagement with epistemically demanding tasks and habituates a pattern of passive surrender rather than active inquiry. The main point here is captured by the commonplace expression "use it or lose it." Like bodybuilding, regular training that progressively increases its intensity is religiously observed because if a trainee neglects this, then he or she will definitely lose his or her muscle gains over time. The same thing can be held for intellectual character building, if students don't engage their cognitive traits in progressively difficult academic challenges, then they might deprive themselves of the growth they deserve, and worse, they might lose gradually the characterological growth they have already gained. What we have here, then, is an instance of active corruption wherein the corruptor actively facilitates the acquisition of vice, in which case, overreliance. To further understand how the passive use of generative AI artifact facilitates epistemic corruption, consider the case of Lard; a student who relies too much on a generative chatbot for writing essays.

Lard is a university student enrolled in a Literature program. He is intelligent and has a propensity for writing, which could have been strategically advantageous for him as a Literature major. Most of the time, his professors would require him to write an essay as a form of assessment. However, instead of writing them, Lard secretly uses a generative chatbot to write the essays for him, claiming them as the product of his own thinking and effort, and carefreely submitting them to his professors. This covert behavior doesn't occur once; rather, it's a behavioral pattern, that is, Lard does this systematically without regard for the kind of academic misconduct he is involved in. Even though Lard has an inkling about the purpose of such academic exercises, that is, to provide the opportunity for him and his fellow students to exercise their independent and creative thinking, he still chooses

to rely on his most-trusted chatbot to do the writing for him because of the convenience and relief it brings.

Lard is clearly on the road towards overreliance, and this is apparent in how he systematically relinquishes his ability to think for himself by letting an external algorithmic system perform the writing process in his stead. In such a case, he has surrendered his cognitive abilities to ChatGPT, instead of using them to think for himself.

I have focused solely on the aspect of writing in which the unreflective practice of algorithmic epistemic outsourcing can interfere with the intellectual character growth of learners but this shouldn't be taken as an indication that such practice is constrained to mere writing per se. Generative AI tools are varied, and so the way in which they facilitate epistemic corruption could also be varied. For example, there are forms of generative AI that is capable of creating complex art and images. Hence, we can preliminarily say that the existence of such algorithmic artifacts opens the possibility that some artists might eventually utilize them to outsource creative labor. And in a similar way, like what has been argued about writing, artists might eventually lose their creativity once they relinquish it to automated systems. Whether if it's the cognitive activity of writing or creating works of art, passive reliance on AI corrupts our intellectual character. And to better understand the strength of this corrupting tendency, we consider, following Kidd (2019), both factors: the psychosocial profile of the agents involved and the normative structures of the educational system in which the practice occurs. The following discussions are at best preliminary, and they can benefit from a more detailed exploration, which can be done in a separate study. However, suffice it to say that these preliminary discussions are sufficient to establish the point I am making.

With respect to the psychological profile of students, several features increase their susceptibility to epistemic corruption. Students are typically under significant academic pressure, with overlapping deadlines and high-performance expectations. According to Krook (2025) "students were more likely to outsource their work to ChatGPT when facing time pressures and workload problems (6017)." With these in mind, we can possibly view David's acquired dependence on AI as a result of academic pressure. Perhaps David was under pressure to submit his academic requirements on time, which is why he was tempted to outsource the writing activities required of him by his professors to AI.

The normative structure of contemporary education further strengthens this corrupting tendency. For the last couple of decades, educational systems have increasingly emphasized measurable outputs, standardized assessment, and performance efficiency (Biesta 2015, 12-13).⁴ Assignments, including writing, are often framed primarily as finished products to be submitted rather than as processes of inquiry. Within such an educational culture, the use of generative AI tools is tacitly incentivized: if students are rewarded for submitting polished work on time, not for demonstrating effort and struggle, then algorithmic epistemic outsourcing becomes a go-to practice. Think of David once again, he deals with academic pressure and high-performance expectations. Given his position, David would naturally leverage available resources that could expedite his pending academic tasks. And since

generative AI tools are almost free and available online, then this provides him the motivation to outsource his tasks to the former.

In light of both discussions, we can then make the conjecture that there is more to the corrupting process of unreflective algorithmic epistemic outsourcing, as there are localized and social factors that are at work. One might even say that the unreflective practice of algorithmic epistemic outsourcing might only be a product of such conditions, and there could be other factors that are not taken into account in this paper. But as far as this paper is concerned, the corrupting tendency still lies and is still attributable to the unreflective practice of algorithmic epistemic outsourcing, whose corrupting tendency may be gauged as moderately strong, especially within educational contexts that prioritize measurable goals that combine with performance pressure.

CONDITIONALITY AND CORRECTIVE CLAIM

Granted, what has been established in the previous section, the corruptive tendency associated with algorithmic epistemic outsourcing should not then be construed as intrinsic to the mere use of generative AI systems. Rather, it should be construed as something that arises under specific conditions that allow or encourage the uncritical delegation of cognitive labor to algorithmic artifacts. In other words, the corruption is unintentional, which in turn means it is conditional; only when the predicament of the users impels them to engage in an unreflective practice of offloading cognitive tasks to algorithmic artifacts. Under these circumstances, students are likely to develop the epistemic vice of algorithmic overreliance, which at the same time diminishes their capacity for independent thinking. Because the corruption arises from these removable conditions, algorithmic epistemic outsourcing is best understood as contingently corrupting rather than intrinsically so. If learners are guided to use AI responsibly and reflectively, and if educational norms support critical engagement rather than passive surrender, then the practice will cease to be corrupting, it can even be utilized to help learners improve their cognitive skills (Suriano et al. 2025, 5).

For this reason, a possible response is not to prohibit AI use but to develop learners' intellectual virtues⁵ like curiosity, intellectual autonomy, open-mindedness and humility,⁶ that would likely enable them to interact with AI more responsibly (Uzskai 2025; Blehm 2024; Naeem 2025). The idea of teaching for intellectual virtues has been widely defended by several scholars and most notably by Jason Baehr (2013). He argues that teaching for intellectual virtues provides a better way of capturing the putative meaning of teaching and learning (Baehr 2013, 253). In the contemporary context, there is a tendency of treating education in a utilitarian way, where it becomes more about qualification⁷ and less about subjectification⁸. Under this paradigm, teaching becomes nothing more than just a technique that revolves around the application of strategies to achieve predetermined outcomes. Correspondingly, learning functions like indoctrination instead of edification that develops students characterologically. Baehr's pedagogical model reorients this tendency. Teaching becomes an intentional process of transforming learners into lovers of epistemic goods,

inculcating in them the relevant motivation and competence as well as judgment that allows them to achieve these goods (Baehr 2013, 253). Learning, in turn, becomes collaborative rather than passive. Students are not simply absorbing content, as if their minds are some sort of memory banks where information is deposited; rather, they are molded into intellectually virtuous individuals. And the latter is possible because a teacher who advocates for growth in intellectual virtues, creates classroom environments that are personal, dialogical, and reflective, in which his or her students are more likely to experience active engagement with the learning materials (Baehr 2013, 254).

We see, therefore, that at the heart of educating for intellectual virtue (EFVA), as developed by Jason Baehr, is a shift from seeing education primarily as the transmission of information or the cultivation of specialized skills, to understanding it as the formation of persons who reliably and responsibly pursue epistemic goods. What is envisioned is not merely what students can know, nor even what they can do, but who they can become as epistemic agents.

Building on this, one might add a further merit, which I anchor on Uzskai's (2024, 23) reflections on the epistemic vulnerability of students to the influence of novel forms of technology. Students are exposed to external conditions and processes that shape their epistemic habits in detrimental ways. That is why, educating for intellectual virtues can be a strategic move as it equips students the proper epistemic motivation, relevant competence and prudential judgment to navigate a distorted technological space without compromising their epistemic agency and integrity.

However, a formidable objection to this proposal is what Clemente (2025) describes as the "Vicious Students Challenge" (112).⁹ The challenge arises from the observation that students are not characterologically neutral when they enter the classroom; rather, they either possess some degree of propensity to be vicious or are already vicious, though no fault of their own (Clemente 2025, 112). In the context of our present inquiry, it might be hypothesized that students may have already developed a high degree of overreliance on genAI. If this is the case, then it might be difficult to educate for virtuous engagement with generative AI tools.

One possible response is to argue that teaching intellectual virtue gradually neutralizes the vicious dispositions already inherent or possessed by the learners. Yet this response raises several difficulties. For example, one might ask: in what sense are vicious dispositions neutralized? Does this mean teaching for open-mindedness will neutralize the students' close-minded tendency? And when we say neutralize, does this mean they become neither vicious nor virtuous? One may further ask: is there a time in the psychological development of an epistemic agent wherein his or her intellectual character is neither vicious nor virtuous? Such claims appear difficult to maintain without substantial empirical and developmental evidence concerning the formation of intellectual character.

Another response would be to admit that teaching for intellectual virtue becomes largely ineffective in such a situation. One might even argue that teaching students who already possess intellectual vices the competences characteristic of virtuous exemplars could unintentionally reinforce rather than weaken or remove their vicious tendencies. Tanesini (2021, 74-95), for example, observes that students who are already disposed to over-own their intellectual limitations, which is the intellectual

vice counterpart of humility, may not be motivated to change even if they are exposed to or are presented with exemplars of humility and pride, as they would view these exemplars as out of reach. And consequently, they will likely maintain and re-affirm their depreciative view of themselves (Tanesini 2021, 74-95). We can paint a similar picture of what it would be like to educate virtues like intellectual autonomy to students who already possess a great degree of overreliance on generative AI tools. In particular, cultivating such virtue may encounter resistance, since over-reliant students may no longer perceive independent intellectual effort as necessary and even valuable. Worse still, attempts to encourage these types of students to practice more reflective and self-directed inquiry may be interpreted by them as intrusive efforts that disrupt practices they already experience as efficient and convenient. Under these circumstances, the pedagogical strategy of educating for intellectual virtue appears, at least initially, inadequate to fully address what Clemente (2024) describes as the “Vicious Students Challenge.”

That is why a third response would involve adopting a complementary approach to educating for intellectual virtue (EFVA), that is, educating against intellectual vice (EAIV) (Clemente 2024). Clemente (2024, 113-115) argues that this complementary approach adopts a comparatively modest and less demanding approach than EFVA. Rather than aiming directly at the cultivation of intellectually virtuous individuals, its primary goal is to help students recognize, resist, and avoid intellectual actions that are characteristic of epistemic vices. Precisely because of this narrower and more modest orientation make it more suitable in contexts wherein students have already developed certain vices. In classroom practice, one way of operationalizing this approach is to resort to what Sullivan and Alfano (2019) describe as ‘negative epistemic exemplars.’ They may refer to literary characters, historical figures, public personalities or even social actors who exhibit intellectually vicious traits. Unlike positive exemplars¹⁰ (see Zagzebski 2017), whose function is aspirational, negative exemplars operate by eliciting negative emotions like agonistic envy (Sullivan and Alfano 2019, 22-25), ambiguity (Sullivan and Alfano 2019, 25-28), or possibly contempt and disgust (Clemente 2024). By exposing students to such exemplars, teachers may elicit negative evaluative and emotional responses toward epistemically vicious qualities and practices, which in turn discourages students from emulating them. Applied to the present discussion, students who already possess a degree of reliance on generative AI may benefit from the ameliorative strategy that identifies and problematizes behaviors characteristic of algorithmic overreliance. And so, using the examples of David, Lard, or other negative exemplars, teachers may meaningfully guide students to critically recognize and thereby avoid cognitive actions characteristic of an over-reliant individual such as passive cognitive outsourcing of vital learning activities like essay writing, diminished reflective engagement with the informational content acquired from external algorithmic tools, and excessive trust to algorithmic outputs. Though this is not to say that EAIV should replace EFVA. Rather, both approaches can function complementarily: one oriented toward the positive cultivation of good qualities of a thinker and the other toward the mitigation and prevention of bad qualities of a thinker.

Nevertheless, even if we consider both pedagogical and ameliorative strategies, and supposedly attain some progress at ameliorating the corruptive effects associated

with algorithmic epistemic outsourcing, such progress will be short-lived as long as the systemic factors remain unaddressed. As Heather Battaly (2022) argues: “systemic changes must be made, if any progress toward virtues is to be lasting (146).”

As argued in the section *on how genAI facilitates epistemic corruption*, two such factors are relevant: first, the psychological predicament of the students, and second, the educational context in which they are situated. Homing in on the latter, we further explored that the practice of algorithmic epistemic outsourcing persists partly because of how it is tacitly incentivized by an educational system that puts a premium on performance, efficiency, and measurable goals (Biesta 2015, 12-13). In light of this, one possible systemic strategy would involve adopting a more humanistic approach to education, one that is concerned with the characterological flourishing of the learners. Such systemic transformation would likely require the collaborative effort among the relevant actors, namely: teachers, school administrators and policy makers that exercise significant authority on shaping the curriculum. And I think the educational proposal of teaching for intellectual virtues is one such humanistic way of viewing education because it understands education not primarily in instrumental terms, but as intrinsically concerned with the formation of good thinkers and responsible learners. Effecting such a paradigm shift would realistically require time but if successful, then the pressures that encourage students to surrender their own cognitive agency to external algorithmic systems may likewise be diminished, since in educational environments that prioritize genuine learning, the need to surrender one’s own thinking to an external system just so that one can measure up to the learning trajectories is no longer incentivized.

Still, this proposal is merely exploratory in nature. Questions concerning its practical implementation, feasibility, and limitations require further investigation. I acknowledge that other systemic changes might also be available beyond what was explored in this paper, though examining these possibilities no longer falls within the scope of this inquiry and therefore remains a task for future investigation.

CONCLUSION

This paper has examined the epistemic and educational challenges posed by the unreflective practice of algorithmic epistemic outsourcing. While it offers undeniable conveniences to student users, it simultaneously puts them at risk of epistemic corruption. By employing Kidd’s corruptionist criticism framework, I have shown that the habitual surrendering of cognitive labor to AI systems can lead to the acquisition of overreliance. However, I have also claimed that this form of epistemic corruption is not intrinsic. Rather, it emerges conditionally as a result of specific psychological and institutional factors. Finally, I have explored possible ways to address the conditional nature of such corrupting process by considering a pedagogical strategy (educating for intellectual virtues), ameliorative strategy (educating against intellectual vices) and a systemic strategy that involves adopting a more humanistic form of education that puts a premium on the individual student’s growth as a thinker, learner, and as a human being.

NOTES

1. This is a term that Ian James Kidd uses as a short hand for describing how certain forms of education facilitate growth in intellectual virtues.

2. Virtue-reliabilism maintains that intellectual virtues are natural faculties that reliably lead an individual to function properly by forming true beliefs and thus, acquiring knowledge. In this proposal, eyesight, memory, introspection, deductive and inductive reasoning are considered as intellectual virtues.

3. Virtue-responsibilism defines intellectual virtues as acquired character traits that involve proper motivation [the love for truth] and competence [virtue-specific skills like listening in the case of open-mindedness]. This proposal considers open-mindedness, autonomy, intellectual humility, and intellectual courage as virtues.

4. These are symptoms of a deeper educational crisis, that cannot be adequately examined within the scope of the present discussion.

5. It is important to point out that in the Educating for Intellectual Virtues literature, intellectual virtues are more predominantly conceived in virtue-responsibilist terms.

6. See Baehr (2021, 34-46) for a discussion of the listed intellectual virtues.

7. I borrowed this term from Biesta (2015). The word qualification is what he uses to describe one of the functions of education. Qualification here means capacitating students with the knowledge and skill set required for a professional job.

8. Another term I borrowed from Biesta (2015), which denotes another function of education. Subjectification refers to the process of transforming learners into independent thinkers, which I believe aligns with the idea of teaching students to become virtuous.

9. Thanks to an anonymous reviewer for urging me to consider objections to strengthen my philosophical argument.

10. See Sullivan and Alfano (2019, 19-21) for a more detailed explanation of the “Admiration-Emulation Model” of epistemic exemplar.

REFERENCES

- Baehr, Jason. 2013. “Educating for Intellectual Virtues: From Theory to Practice.” *Journal of Philosophy of Education* 47 (2): 248–62. <https://doi.org/10.1111/1467-9752.12023>.
- Baehr, Jason. 2021. “The Structure of Intellectual Vices.” In *Vice Epistemology*, edited by I. J. Kidd, Heather Battaly, and Quassim Cassam. Routledge.
- Battaly, Heather. 2023. “Intellectual Character Education: Some Lessons from Vice Epistemology.” In *Handbook of Philosophy of Education*, edited by R. Curren. Routledge.
- Biesta, Gert J. 2015. *Good Education in an Age of Measurement: Ethics, Politics, Democracy*. Routledge.
- Cassam, Quassim. 2019. *Vices of the Mind: From the Intellectual to the Political*. Oxford University Press.

- Clemente, Noel. 2024. "Educating Against Intellectual Vices." *Ethics and Education* 19 (1): 109-23. <https://doi.org/10.1080/17449642.2024.2335133>
- Flynn, James R. 2012. *Are We Getting Smarter?: Rising IQ in the Twenty-First Century*. Cambridge University Press.
- Flynn, James R. 2016. *Does Your Family Make You Smarter?: Nature, Nurture, and Human Autonomy*. Cambridge University Press.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Kidd, Ian J. 2019. "Epistemic Corruption and Education." *Episteme* 16 (2): 220–35. <https://doi.org/10.1017/epi.2018.3>.
- Kidd, Ian J. 2020. "Epistemic Corruption and Social Oppression." In *Vice Epistemology*, edited by I. J. Kidd, Heather Battaly, and Quassim Cassam. London: Routledge.
- Krook, Julia. 2025. "When Autonomy Breaks: The Hidden Existential Risk of AI." *AI & Society* 40 (8): 6011–24. <https://doi.org/10.1007/s00146-025-02397-5>.
- Medina, Jose. 2012. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford University Press.
- Naeem, Hafsa. 2025. "Teaching Skills and Intellectual Virtues with Generative AI." *Episteme*: 1–18. <https://doi.org/10.1017/epi.2025.10089>.
- OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774. <https://arxiv.org/pdf/2303.08774.pdf>.
- Pritchard, Duncan H. 2016. "Intellectual Virtue, Extended Cognition, and the Epistemology of Education." In *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, edited by J. Baehr. Routledge.
- Shaw, S. D., & Nave, G. 2026. Thinking Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender." OSF Preprints. https://doi.org/10.31234/osf.io/yk25n_v1
- Tanesini, Alessandra. 2021. *The Mismeasure of the Self: A Study in Vice Epistemology*. Oxford University Press.
- Suriano, R., Plebe, A., Acciai, A., & Fabio, R. A. 2024. Student Interaction with ChatGPT can Promote Complex Critical Thinking Skills. *Learning and Instruction*, 95 (2025), 12011. <https://doi.org/10.1016/j.learninstruc.2024.102011>
- Uszkai, Radu. 2024. "Fostering Intellectual and Ethical Virtues in the Age of Artificial Intelligence: The Need for Educators-in-the-Loop." In *School Children and the Challenge of Managing AI Technologies: Fostering a Critical Relationship through Aesthetic Experiences*, edited by E. Guarcello and A. Longo. Routledge.
- Zhai, C., S. Wibowo, and L. D. Li. 2024. "The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review." *Smart Learning Environments* 11 (28): 1–37. <https://doi.org/10.1186/s40561-024-00316-7>.