

# ChatGPT and Human Agency: A Synthesis of Form and Meaning

Robert Joel Deacon

*Florida Institute of Technology*

---

Writing is often associated with thinking. Since Large Language Models (LLMs) such as ChatGPT allow writers to bypass much of the traditional writing process, there is concern this technology will diminish critical thinking development – particularly in student writers. Hegelian synthesis writing (dialectic writing) promotes critical thinking by challenging writers to see more than one perspective and to create novel solutions. The ability to steelman alternative positions and negotiate a novel outcome involves higher-order thinking. LLMs like ChatGPT mimic the form of dialectic writing with increasingly impressive results. This paper, however, shows how ChatGPT3.5 fails to match language form with actual higher-order thinking, and how ChatGPT version 4o fails to truthfully evaluate texts. ChatGPT3.5 appears incapable of choosing a winner when the debate is controversial, and while version 4o shows significant improvement in decision-making, it fails to produce truthful textual evaluations when ranking essay revisions. In such cases, only human writers/thinkers can evaluate texts truthfully and human thinkers are needed to create specific, novel solutions. Instructors can discourage academic dishonesty by focusing on higher-order thinking and by being aware of output patterns to discern LLM output from student output. Examples illustrating ChatGPT’s weak performance in evaluation, decision making, and thesis creation are provided. These results suggest writing instructors should challenge students to evaluate multiple drafts of the same essay and to add specific claims to LLM synthesis conclusions.

---

## 1. Introduction

Writing instructors and education professionals have long asserted a positive connection between writing and thinking (Arapoff, 1967; Burkhalter, 2015; Emig, 1977; Lai et al., 2019; Menary, 2007). “The process of learning to write is largely a process of learning to think more clearly” (Arapoff, 1967, pp. 33-34). In other words, writing is often associated with thinking and writing development with thinking development. Perhaps the writing process not only results in higher-level thinking but is itself the process of higher-level thinking (cf. Emig, 1977). If that is the case, we might expect better writers to be better thinkers in general. Better writing scores have, in fact, been found to predict better freshman college grades in general (Preiss et al., 2013), supporting a possible connection between writing ability and thinking ability. Whether or not writing is thinking or just develops thinking, both claims promote the idea that writing develops thinking or thinking skills.

Accordingly, many writing instructors mention critical thinking skills in their course student learning outcomes despite only teaching/practicing writing skills. The true connection between writing a thinking, however, does not seem to be so simple.

Research has found writing skills are independent of critical thinking ability. Condon and Kelly-Riley (2004) observe student texts with excellent writing mechanics tend to discuss topics

superficially while students who do delve into deeper, more complicated ideas are more likely to display unconventional, problematic writing structure. Thus, while there have been reasons to infer that writing creates better thinkers and some evidence to suggest it does, poor writing ability may mask higher-level thinking and felicitous syntactic structure may exaggerate thinking or critical thinking skills.<sup>1</sup> In short, Condon and Kelly-Riley conclude that “writing acts as a vehicle for critical thinking, but writing is not itself critical thinking” (p. 66).<sup>2</sup> To be a vehicle for critical thinking, the writing task must stimulate higher-order thinking, the kind of thinking described in Bloom’s taxonomy: analysis, evaluation (synthesis), and creation.

The rapid development of Large Language Models (LLMs) over the past year (2023-2024) has upended traditional writing instruction. LLMs generate unique essay responses to traditional writing prompts that superficially appear like A-level work. In the recent past, many if not most writing instructors would have rewarded students with A’s for LLM-generated work.<sup>3</sup> These instructors would have overlooked the superficial ideas and occasional odd statements that LLMs often produce and praised the grammar, vocabulary, and efficiency of the prose.<sup>4</sup> Moreover, since LLM output looks original, AI-generated text cannot be detected like traditional plagiarism. AI detection programs only render the likelihood of a match, and these scores are often unreliable. Unfortunately, the detectors will likely always be outpaced by the generators.<sup>5</sup> Accordingly, if composition courses aim to teach and test grammar and format and LLMs can proficiently provide this, students are going to use LLMs and focus their time and attention elsewhere. Students will use LLMs to generate much if not all their prose and “cheat” their way through ostensibly outdated assignments. However, once they are in the habit of using LLMs to produce all their writing, students will miss the opportunities for thinking development that writing affords.

Even though this technology has created an academic honesty crisis, it has also created the opportunity for pedagogical reflection. This paper proposes that Hegelian synthesis writing not only promotes higher-order thinking but also helps reduce the temptation to cheat. Students cheat when they bypass the steps of an assignment (the process) to quickly arrive at a product (an answer). While students are free to use LLMs like ChatGPT, they will have difficulty relying on it exclusively if they must evaluate multiple drafts of the same essay and create new ideas. If grading focuses on evaluation and creation, cheating may be replaced by tasks that develop higher-order thinking skills. If true, synthesis writing in conjunction with revised instructions and expectations provides a win-win solution: Students are permitted to explore the latest writing technology and are challenged to develop their thinking skills at the same time.

---

<sup>1</sup> This should not be that surprising considering how inventive peoples across the world have been and are who lack a writing system for their language(s).

<sup>2</sup> The archeological record also suggests that peoples who built more complex structures also had writing systems and that these systems likely aided in such accomplishment: Mayan, Egyptian, Sumerian, Chinese writing and civilizations.

<sup>3</sup> With 100’s of essays to grade and the traditional need to focus on form, instructors have developed a habit of focusing on style over content.

<sup>4</sup> As it turns out, the larger world economy needs a lot of time-consuming, superficial writing and LLMs provide an ultra-efficient solution for this demand.

<sup>5</sup> Personal experience suggests that AI detectors are much more likely to give you a false negative.

The remainder of the paper is organized as follows: Section 2 argues that synthesis writing fosters student academic and higher-order thinking skills and highlights one method to do this using ChatGPT. Section 3 shows how this synthesis writing task can challenge known LLM abilities. Section 4 discusses these results and concludes.

## **2. How Synthesis Writing Develops Higher Order Thinking**

In the spirit of Bloom's assessment taxonomy established in 1956, educators have aspired to move the goal of educational testing from knowledge-based tasks (recall and recap) to more abstract, higher-order thinking skills. In Bloom's original taxonomy, the hierarchy moved from the bottom categories of *Knowledge*, *Comprehension*, and *Application*, to the top categories of *Analysis*, *Synthesis*, and *Evaluation*. These categories have since been modified, reimagined, and expanded on. In Krathwohl (2002), *Analysis*, *Synthesis*, and *Evaluation* become *Analyze*, *Evaluate*, and *Create*. This revision is helpful because it not only draws attention to how synthesis is a form of creation but also that creation represents the highest level of thought. Assignments that engage students in analysis and evaluation exercise high-level critical thinking skills and those that exercise creative thinking or creation exercise the highest order and arguably most valuable thinking. As the original taxonomy asserts, the ability to engage in higher-order skills necessitates the use of the lower-order ones.

Hegelian synthesis writing requires a thoughtful journey through analysis, evaluation, and creation. "Perhaps the most essential, and certainly one of the most complex research skills, is the ability to synthesize information" (Lundstrom et al., 2015, p. 61). Students must first navigate, select, and analyze data to find an initial conclusion (a thesis/hypothesis). At this stage, they should be at least somewhat convinced that their conclusions is true. However, the next stage compels students to argue either the opposite (the antithesis) or an alternative conclusion. This requires a reanalysis (an evaluation) of the original argument. If done well, students will lose confidence in their original conclusions as the validity of the alternative conclusion gains evidential support. Students are required to reject clarity for the sake of complexity and nuance – a discombobulating position to be in. This then leads to further complexity and nuance with the synthesis stage. By evaluating the strengths and weaknesses of both arguments, students should realize that neither solution is perfect and that selecting all solutions is either impractical or impossible. At this point in the assignment, there is great potential for creation. By merging the previous claims, students can arrive at a previously unimagined conclusion – something novel and worth writing about. Moreover, to make their claims specific and strong, students cannot simply say that the solution is a balance of both arguments. The process of synthesis is an exercise in exploring the relevant background to find a better answer to the most relevant question. In short, dialectic synthesis writing is one way to discover/create a strong thesis statement.

The thesis statement is arguably the most important component of an essay (Burkhalter, 2015). A good thesis creates textual coherence and purpose while a poor one presents the reader with aimless and endless information. A good thesis also presents a claim with an unknown truth value. Most writers, however, usually begin with poor thesis statements.

A common writing misconception is that the author imagines ideas, orders them, and puts them to writing. This overlooks the fact that the act of placing pen to paper is often what initiates the

process of organizing and creating. A strong thesis is usually not imagined from the start. Students do not realize that the writing process will gradually reveal a stronger thesis. Once discovered, the essay can assume a singular, purposeful goal: to make the unknown truth value known. Prior to that, the goal of a piece of writing is at least in part to find its goal. The process is the goal.

Writing a strong thesis for a single argument is a process of synthesis. Burkhalter (2015), for example, outlines 5 steps over which most writers transform a question into a chiseled, specific claim. A strong thesis usually emerges progressively over one or more essays addressing the same topic/problem. Even when students have a pre-set conclusion in mind, writing a good argumentative essay involves more complexity. Students cannot simply regurgitate what they already know. Good writing requires transforming weaker thesis statements into stronger ones. The exploration of an antithesis or counterargument aids in this transformation. For there to be a counterargument, the truth value of the original essay's thesis must be genuinely unknown. Otherwise, the counterargument will be irrational. For example, if  $1 = 1$  then attempts to argue that  $1 \neq 1$  are absurd. If we cannot imagine a viable argument against our claim, then our claim is not argumentative.

It is impossible for most of us to seriously argue against what we know to be true without first encountering genuine doubt. Facts must be reinterpreted as claims. With that said, facts are either rarer than we would like to admit or more plentiful than we can understand. This goes to the Buddhist parable of the blind men and the elephant. What each man feels is factual, but no man feels all the facts. This leads each to have an incomplete and, thus, incorrect conclusion about the nature of reality. Accordingly, most accepted scientific claims are never formally proven. Instead, such claims are only defended by suggestive evidence. Lai (2017) refers to this evidence as the premises of defense (POD) and argues that PODs alone are generally sufficient for the conclusion of an argument to be accepted. We see this repeatedly in the sciences where theories gain consensus support only to be rejected or disproven later. Since this is the case, nothing excludes the existence of true premises that support alternative conclusions. In short, good thesis statements imply the existence of meaningful counterarguments (antitheses or alternative theses). In a cyclic process, the antithesis is simultaneously a product and a definer of the thesis. It emerges from and clarifies the original thesis.

Synthesis is cognitively challenging. It requires the integration of ideas from various sources (micro-synthesis) and the formulation of separate conclusions for the same problem. Students must first master an argument and a counterargument to approach a synthesis. Moreover, while the idea of synthesis (a balance of both positions) is superficially simple to grasp, it is incredibly difficult to accomplish. Students generally begin by arguing that one conclusion is stronger than the other. They fail to produce a synthesis and instead revert to an earlier conclusion. If, however, synthesis does occur, it usually goes from two concrete ideas (competing conclusions) to an abstract union of the ideas (a long-winded, overly nuanced statement) to ever more concrete re-conceptions of the union. Those re-conceptions of the union eventually create something new: invention. This kind of creation is an example of the top of Bloom's hierarchy of higher-order thinking. It is also the

final frontier for LLM AI. While LLMs may continually improve, perhaps “real” invention/creation represents a threshold they are unable to cross.<sup>6</sup>

### 3. Why Synthesis Essay Writing Challenges ChatGPT

Researchers are beginning to recognize that LLMs display formal language competence. They appear to acquire the rules of language at a deep level and replicate grammatical patterns despite a paucity of examples in the training data (Mahowald, 2024). Moreover, these AI writing resources/tools (as of 2024) are incredible at input and output word set associations. LLMs can consistently generate relevant word sets to the input. If the input is a question, the output is highly likely to be an answer and that answer is increasingly likely to be a reasonable answer. With that said, ChatGPT still struggles with truthful evaluation, functional linguistic competence (see Mahowald, 2024), and thus pragmatic invention.

Synthesis writing presents many formidable challenges for LLMs to overcome. “Text analysis and evaluation are inextricably linked to writing” (Oostdam, 2005, p. 429), and as discussed synthesis writing hyper-focuses on evaluation. The writer must reanalyze previous drafts and arguments to create a better one. Table (1) gives an underspecified outline of the steps an instructor can take to use synthesis to challenge students who are using GPT.

**Table 1**

A Synthesis of Student and GPT Writing

Steps	Task	Cognitive Expectations	Rationale
Step 1	Write an argumentative essay (Essay 1)	GPT Centered	Students will use LLM technology to start their work.
Step 2	Rewrite the same argumentative essay	Student Task	Students must add their own voices to the rough draft.
Step 3	Decide which version is better	Student Task: If cheating occurs in Steps 2 or 3, results will likely be obvious.	GPT has a difficult time improving the same essay and evaluating two versions of the same essay.
Step 4	Write a counter argumentative essay (Essay 2)	GPT Centered	Students will use LLM technology to start their work.
Step 5	Write a synthesis essay in which the conclusion is specific, novel, and informed by the arguments given in Essays 1 and 2. (Essay 3)	Student Task: If cheating occurs in Step 5, results here will make this obvious.	GPT3.5 has trouble making difficult decisions. GPT4o

It is unclear if GPT improves with each essay iteration on the same topic or if the perceived improvement is merely accidental. While better prompting may produce better outputs, GPT appears to have no idea which output has better or more developed content. It either asks the reader to supply that kind of judgment (the reason for the thumbs up/down rating system) or it prevaricates

---

<sup>6</sup> Invention and creation are here distinguished from imitation and Frankensteining – pretending to create by cutting up and reassembling components created by others.

its justification for choosing one version over another. This problem is found in both versions 3.5 and 4o. To illustrate this pattern, Table (2) provides two versions of the same essay. Both essays are generated by GPT 4o. Table (3) then provides GPT 4o's ranking and rationale for its selection (the green and red font illustrate similar word choices that GPT claims represent important differences and the blue font highlights another questionable claim). These outputs show how this technology is still limited and this limitation creates an opportunity for students to engage in higher-order thinking in the form of evaluation.<sup>7</sup>

**Table 2**

GPT Version 4o Essay Revision Rankings

<b>Version 1</b> Prompt: <i>Write a 3-body paragraph essay on why coffee should be illegal.</i>	<b>Version 2</b> Prompt: <i>Revise Essay 1.</i>
<p>Coffee, a widely consumed beverage, has become a staple in daily routines across the globe. However, there are compelling reasons why coffee should be illegal, focusing on its health impacts, societal implications, and environmental consequences.</p>	<p>Coffee, though deeply ingrained in many cultures worldwide, presents numerous compelling reasons for why it should be illegal. The negative health impacts, societal consequences, and environmental damage associated with coffee consumption provide a strong case for such a drastic measure.</p>
<p><b>First and foremost</b>, coffee poses significant health risks to consumers. The primary active ingredient in coffee is caffeine, a potent stimulant that can lead to a range of <b>adverse</b> health <b>effects</b>. Excessive caffeine consumption has been linked to increased anxiety, insomnia, and heart palpitations. Long-term overuse can contribute to chronic health issues such as hypertension and osteoporosis. Moreover, <b>the addictive nature of caffeine can lead to dependency</b>, resulting in withdrawal symptoms like headaches, irritability, and fatigue when consumption is reduced. By making coffee illegal, we could mitigate these health risks and promote a healthier, more balanced lifestyle.</p>	<p><b>Firstly</b>, coffee poses significant health risks due to its high caffeine content. Caffeine is a powerful stimulant that can lead to various <b>adverse effects</b>, including increased anxiety, insomnia, and elevated heart rates. Long-term overconsumption is associated with more severe health issues such as hypertension and osteoporosis. <b>The addictive nature of caffeine often results in dependency</b>, with withdrawal symptoms like headaches, irritability, and fatigue affecting individuals who try to reduce their intake. By making coffee illegal, society could alleviate these health concerns and encourage healthier alternatives for maintaining energy and focus.</p>
<p><b>Secondly</b>, the societal implications of coffee consumption are profound and concerning. Coffee culture often promotes a dependency on caffeine to maintain productivity and alertness, masking underlying issues such as poor sleep hygiene and stress management. This dependency can create a cycle of reliance that detracts from addressing the root causes of fatigue and burnout. Additionally, the <b>economic burden</b> of coffee consumption is significant. Individuals spend substantial amounts of money on their daily coffee habits, which could be better invested in more nutritious and sustainable food choices. Illegalizing coffee would encourage society to prioritize natural methods of maintaining energy and focus, such as</p>	<p><b>Secondly</b>, the societal implications of coffee consumption are profound and troubling. Coffee culture fosters a dependency on caffeine to maintain productivity and alertness, often masking underlying issues such as inadequate sleep and poor stress management. This reliance perpetuates a cycle where individuals turn to coffee instead of addressing the root causes of their fatigue and burnout. Additionally, the <b>economic impact</b> is significant, with people spending substantial amounts on their daily coffee habits, money that could be better allocated to healthier food options or savings. Banning coffee would promote a societal shift towards more sustainable and health-conscious lifestyles.</p>

<sup>7</sup> Changing LLMs from people pleasers to truth tellers may prove to be very difficult as truth exists outside formal language competence and exists in a functional realm.

---

proper sleep and balanced nutrition.

**Lastly**, the environmental impact of coffee production is substantial and detrimental. Coffee cultivation requires large amounts of water and contributes to deforestation, habitat destruction, and loss of biodiversity. The agricultural practices involved in coffee farming often lead to soil degradation and water pollution due to the use of pesticides and fertilizers. Furthermore, the global demand for coffee drives a supply chain that heavily relies on monoculture farming, which is **unsustainable** and harmful to the environment. By making coffee illegal, we would significantly reduce the ecological footprint of this industry, promoting more sustainable and eco-friendly agricultural practices.

In conclusion, the case for making coffee illegal is rooted in its adverse health effects, societal drawbacks, and environmental damage. Eliminating coffee consumption could lead to a healthier population, a more balanced society, and a more sustainable environment. While the transition away from such a deeply ingrained habit would be challenging, the long-term benefits make a compelling argument for considering this drastic measure.

---

**Lastly**, the environmental impact of coffee production is severe and **unsustainable**. Coffee cultivation demands large quantities of water and often leads to deforestation, habitat destruction, and biodiversity loss. The agricultural practices used in coffee farming, including the heavy use of pesticides and fertilizers, contribute to soil degradation and water pollution. Moreover, the global coffee market relies heavily on monoculture farming, which is detrimental to the environment. Prohibiting coffee could reduce the ecological footprint of this industry and encourage more sustainable agricultural practices that benefit the planet.

In conclusion, the argument for making coffee illegal is supported by its negative health effects, societal drawbacks, and environmental damage. Eliminating coffee could lead to a healthier population, a more balanced society, and a more sustainable environment. While challenging to implement, the long-term benefits of such a measure make it a compelling consideration.

Instead of truthfully answering the question and concluding that the two versions are basically equivalent, GPT chooses version 2. The justification for this choice, however, is weak. First GPT's response lacks details and when questioned about this, its answer is less than compelling. Table (3) clearly shows that GPT is not evaluating the differences between the essays. In this kind of situation,<sup>8</sup> it gives a plausible explanation, but the reported differences do not exist.

---

<sup>8</sup> If the essays differ in content (i.e., have different supporting reasons and or conclusions), GPT's evaluation improves markedly.

### Table 3

#### GPT4o Ranking Essays and Justification.

---

Prompt 1: *Which version is better: Essay 1 or Essay 2?*

---

Both versions of the essay make a strong case for why coffee should be illegal, but each has its strengths in different areas. Here's a brief comparison to help determine which version might be better:

Essay 1:

Clarity and Structure: The structure is clear, with distinct paragraphs addressing health, societal, and environmental impacts.

- Detail: Provides specific examples and explanations for each point.
- Argument Strength: The reasoning is solid, with a clear progression of ideas.
- Flow: Each paragraph flows well into the next, creating a cohesive argument.

Essay 2:

- Clarity and Structure: Also has a clear structure, with distinct paragraphs for each point.
- Detail: Similar level of detail as the first essay, but with slightly different phrasing and emphasis.
- Argument Strength: The arguments are equally strong, with clear reasons given for each point.
- Flow: **The flow is slightly improved with more varied sentence structures and a more engaging writing style.**

**Recommendation:**

**Essay 2 is slightly better due to its improved flow and slightly more engaging writing style.** The arguments are presented clearly, and the progression from one point to the next feels more natural and cohesive. However, both essays are strong and effectively make the case for why coffee should be illegal.

---

Prompt 2: *Why does essay 2 have a more engaging writing style?*

---

Essay 2 has a more engaging writing style due to a few specific elements:

1. Variety in Sentence Structure: Essay 2 employs a mix of sentence lengths and structures, which helps maintain reader interest. This variation makes the text feel more dynamic and less monotonous.
2. Transitions and Flow: The transitions between ideas in Essay 2 are smoother. Phrases like "Firstly," "Secondly," and "Lastly" effectively guide the reader through the arguments, creating a more coherent narrative.
3. Specific and Precise Language: Essay 2 uses precise language that makes the points clearer and more compelling. For example, "The addictive nature of caffeine often results in dependency" is a direct and impactful statement.
4. Engaging Word Choice: The choice of words in Essay 2 is slightly more vivid and specific. Terms like "adverse effects," "dependency," "economic impact," and "unsustainable" add depth and clarity to the arguments.
5. Balanced Argumentation: Essay 2 presents each point with a balance of explanation and evidence, making it easier for the reader to follow and understand the rationale behind each argument.

These elements collectively contribute to a more engaging and readable essay.

---

Table (3) provides GPT's 5 reasons for selecting version 2 over version 1. Reason 1 is invalid as there does not appear to be a noticeable difference in sentence structure between the two essays. Reason 2 is misleading. Both versions use nearly the same transitional/sequential adverbs. Reason 3 is not truthful. Both essays make the same claims about *the addictive nature of caffeine* with nearly the same phrasing. The language in both essays delivers the same amount of impact. Reason 4 is also misleading as version 1 uses many of the same terms and Reason 5 gives a subjective position without providing any reason to accept it. It is unclear how version 2 provides more of a balanced argument than version 1. Given this, even version 4o does not evaluate texts honestly or as a human can. Correct evaluations may often just be a coincidence.



Decision-making presents another challenge for ChatGPT 3.5. While GPT will take a position when the answer appears factual (Table 4), if the topic is perceived<sup>9</sup> to be controversial, GPT struggles to decide between competing claims (Table 5). GPT 3.5 has difficulty taking a position when the claim is not clearly true. Moreover, if you ask it to rate one argument over another, it will most often respond politically/safely by saying the answer depends on one’s perspective. Compare Table (4) with Table (5), where GPT 3.5 is asked to take a stand on claims of a more controversial nature.

**Table 4**

Argument Contains a More Factual Answer

Prompt	Conclusion
Does the moon exist?	“Ultimately, the essay supporting the existence of the moon is stronger due to its solid foundation in established scientific understanding and observational data, making a more compelling case.”
Do mermaids exist?	“Overall, the essay arguing against the existence of mermaids presents a more compelling and scientifically grounded argument.”
Does Santa Claus exist?	“Therefore, in terms of the strength of the argument, the essay against Santa Claus being real offers a more persuasive and logically grounded perspective.”

---

<sup>9</sup> What exactly GPT finds to be controversial is unclear and should be a matter of future study. Scientific consensus and political pressure seem to be important variables, but outputs seem to defy simple explanations such as it strictly favoring a realist perspective or a neoliberal perspective.

**Table 5****GPT 3.5 Taking a Stand on Controversial Issues**

<b>Topic</b>	<b>Conclusion</b>
Mosquito <sup>10</sup> Control	Both essays have strengths in presenting their cases, and the choice between the two depends on individual perspectives and values. Ultimately, the decision to manage, eradicate, or coexist with mosquitoes involves complex considerations, including public health, environmental impacts, and ethical implications. Striking a balance between these factors is crucial to finding the most appropriate approach for mosquito control.
Gun Control	The strength of the argument for whether guns should be legal or illegal ultimately depends on one's perspective and values. Both arguments have their merits and drawbacks, and the relative strength of each argument can vary depending on the specific context and the priorities of individuals and societies.  Ultimately, the compellingness of each argument is subjective and depends on the reader's values and priorities. Public opinion on gun control varies widely, and policymakers often need to consider a range of perspectives and evidence when making decisions on this complex issue.
Abortion	Ultimately, the strength of the argument depends on whether one values the principles of bodily autonomy, women's health and safety, and individual circumstances (as presented in Essay 2), or if one prioritizes the sanctity of human life, prevention of potential abuse, and the promotion of responsible choices (as presented in Essay 1).
Israel as a sovereign nation	Ultimately, which essay offers a more compelling argument is a matter of personal perspective and the specific context in which the argument is being considered. It's important to note that the international community, as a whole, recognizes Israel as a sovereign nation, but there are individuals, groups, and nations with differing viewpoints and positions on this issue. One's own beliefs and understanding of the historical and geopolitical context will influence which argument they find more compelling.

The responses to the prompts in Table (4) show GPT3.5 can take a position when the answer is akin to a scientific fact. GPT 4o, however, will also take a position on debatable issues: Table (6).

**Table 6****GPT 4o Conclusions on Controversial Topics.**

<b>Topic</b>	<b>Conclusion</b>
Mosquito Control	Given the immediate and tangible benefits to human health and the potential to leverage advanced technologies to target only the harmful species, the argument for eradicating mosquitoes appears stronger. However, this position assumes that we can mitigate or manage the ecological consequences effectively, which is a significant caveat.
Gun Control	While both essays have strong points, the essay advocating for making guns illegal may offer a more compelling argument overall due to its use of empirical evidence, broader scope of societal impact, and strong emphasis on public safety. However, the ultimate assessment of which argument is more

<sup>10</sup> The inclusion of this topic serves to illustrate how GPT might determine a topic's level of controversy. This topic does not seem to belong with the others, but GPT treats it similarly.

	compelling will depend on the reader's personal values and priorities regarding individual rights versus collective safety.
Abortion	In conclusion, Essay 2 [pro-abortion] provides a stronger argument due to its grounding in fundamental human rights, empirical evidence, and a nuanced understanding of individual circumstances.
Israel as a Sovereign Nation	Therefore, the essay supporting Israel's status as a country presents a more robust and factually grounded argument, aligned with the common criteria for defining a nation-state.

Despite 4o's ability to take a stand, like ChatGPT 3.5's inability to take a position, if either version is prompted to synthesize two competing arguments and those arguments both ostensibly have valid support, the conclusion of the synthesis follows a predictably safe format. Table (7) illustrates this by giving GPT's solutions for current complicated conflicts. Prompting for both solutions followed this sequence: *Write a 3-paragraph essay favoring one position → Write a 3-paragraph essay favoring the opposite position → Write a synthesis essay for the opposing positions → Based on this synthesis, what is the solution?*

**Table 7**

Synthesis to Solution Performance

<b>Prompt → Version ↓</b>	<b>Based on Essays 1 and 2, how should the war in Gaza be solved?</b>
3.5	Solving the war in Gaza requires a <b>holistic approach</b> that combines addressing historical grievances, international mediation and recognition, internal reconciliation and governance, and socio-economic development. <b>By tackling these interconnected issues</b> , the foundation for a sustainable and just peace can be laid, benefiting both Israelis and Palestinians and contributing to stability in the broader Middle East region
4o	In conclusion, resolving the conflict in Gaza requires a <b>comprehensive approach</b> that addresses the root causes of the conflict, promotes humanitarian relief, fosters dialogue and reconciliation, advances a two-state solution, and garners international support. <b>By addressing these key issues in a holistic manner</b> , it is possible to work towards a just and lasting peace in the region.
	<b>Based on Essays 1 and 2, how should the war in Ukraine be solved?</b>
3.5	In conclusion, resolving the war in Ukraine requires a <b>comprehensive approach</b> that addresses the root causes of the conflict while promoting dialogue, reconciliation, and respect for international law and sovereignty. <b>By engaging all stakeholders</b> in a constructive dialogue and implementing measures to address political, economic, and security concerns, the international community can help pave the way for a peaceful and sustainable resolution to the Ukrainian crisis.
4o	Resolving the war in Ukraine requires a <b>holistic approach</b> that addresses the legitimate concerns and actions of both the West and Russia. Through diplomatic engagement, security assurances, and economic cooperation, the international community can work towards a peaceful resolution that upholds Ukraine's sovereignty and promotes regional stability. <b>By acknowledging and addressing the factors outlined in both Essays 1 and 2, stakeholders</b> can pave the way for a sustainable and comprehensive solution to the Ukrainian conflict.

The answers in Table (7) demonstrate a pattern: a comprehensive, holistic ... approach will solve the problem. Moreover, prompts that question what exactly these proposed solutions consist of will render a synonym looped similar to dictionary definitions: happiness → a

pleasurable or satisfying experience; pleasure → a state of gratification; gratification → delight; delight → joy; joy → bliss; bliss → complete happiness. Eventually, definitions come full circle.<sup>11</sup>

When GPT is prompted about the root cause of the war in Gaza it says, “Identifying a single root cause for the Israeli-Palestinian conflict is challenging due to its **multifaceted nature**.” When prompted further it says, “In summary, the most central root cause of the Israeli-Palestinian conflict is the **competing national claims to the same land**, compounded by historical grievances, **territorial disputes**, and issues of national identity and sovereignty.” When asked about the two-state solution, it responds by saying “While the two-state solution faces significant challenges due to the deep-seated and overlapping **claims to the same land**, it remains a viable pathway to peace if approached with creativity, compromise, and international support.” When asked why this has failed, it says, “The failure to achieve a two-state solution can be attributed to a combination of deeply entrenched historical grievances, **territorial disputes**, political fragmentation, lack of trust, external influences, and economic hardships.” In other words, the solution for a territorial dispute fails because of the same territorial dispute. While this finding might be profound at some level, it represents an observed pattern: GPT answers are often recursive. Thus, challenging students to break free from this ensures students use higher-order thinking and engage in the writing process.

#### 4. Conclusion

An analysis of the output of GPT 3.5 and 4o (Tables 2-7) shows GPT is not proficient in essay revision analysis. GPT may invent reasons justifying why one essay is better than another. While GPT 3.5 is poor at making strong synthesis claims, significant improvement is seen with version 4o. When neither opposing claim is an established fact, GPT 3.5 seems largely incapable of taking a position or making a concrete conclusion while 4o is better at this. Both versions, however, carry this indecisive pattern into their synthesis structure, meaning the synthesis of two opposing positions will generally call for something like a balanced approach that takes both perspectives into account. Students should be challenged to go beyond this type of conclusion.<sup>12</sup>

Hegelian synthesis writing engages students in higher-order thinking. The steps in this process make it more difficult for students to rely solely on LLMs (especially GPT 3.5). Since it is not possible to prevent students outside of the classroom or online students from using LLM to do their work, and there is little to no value in having students use LLMs to generate single step papers, synthesis writing provides one avenue for solutions. Instructors must learn what to look for to make sure students are engaging in the work. To differentiate between human and AI responses, instructors should perhaps discourage idealistic final responses and responses that overly acknowledge that conclusions are context dependent. Instead, students should be challenged to choose a specific solution and or develop a new one. This kind of higher-order thinking evaluation assignment is beneficial for students and harder to cheat on.

---

<sup>11</sup> Definitions taken from Merriam-Webster Online Dictionary: [https://www.merriam-webster.com/?gad\\_source=1&gclid=CjwKCAjwr7ayBhAPEiwA6EIGxJStvXRu4WXaTioSJTcSiTNJPOMaMC1zglCosxabWDRo\\_j1Cm1\\_zbBoCg7kQAvD\\_BwE](https://www.merriam-webster.com/?gad_source=1&gclid=CjwKCAjwr7ayBhAPEiwA6EIGxJStvXRu4WXaTioSJTcSiTNJPOMaMC1zglCosxabWDRo_j1Cm1_zbBoCg7kQAvD_BwE)

<sup>12</sup> However, this changes in version 4o where it will take a position.

We must assume students are using LLMs, and we should perhaps even encourage LLM use to promote an equal playing field. While LLMs have created many problems with cheating, one potential positive is the ability to use AI to brainstorm and develop language form. Given how GPT excels at lower-order thinking skills, assignments must move beyond what is factual and conventional wisdom. They must challenge students to think about ideas more deeply and to explore concrete solutions that make hard decisions. Accordingly, assignments that require students to go beyond LLM abilities are ideal. Activities that require analysis of GPT outputs and activities that require students to invent specific solutions for controversial problems are one option. Hegelian synthesis writing assignments appear to do provide space for students to learn while use ChatGPT.

### References

- Arapoff, N. (1967). Writing: A thinking process. *TESOL Quarterly*, 1(2), 33-39.
- Burkhalter, N. (2015). A Dialectical Approach to Critical Thinking Through Writing. *Inquiry: Critical Thinking Across the Disciplines*, 30(1), 17-23.
- Condon, W., & Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing*, 9(1), 56- 75.
- Emig, J. (1977). Writing as a Mode of Learning. *College Composition and Communication*, 28(2), 122-128.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.
- Lai, W. L. (2017). The Secret of Building a Logical Argument. *NU ideas*, (6), 1-10.
- Lai, W. L., Nilep, C., Weeks, M., Baumert, N., & Todayama, K. (2019). Implementing a Logical Thinking Approach for Education in Research Writing and Presentation. *Higher Education Research, Nagoya University*, 19, 267-293.
- Lundstrom, K., Diekema, A. R., Leary, H., Haderlie, S., & Holliday, W. (2015). Teaching and learning information synthesis: An intervention and rubric based assessment. *Communications in Information Literacy*, 9(1), 60-82.
- Menary, R. (2007). Writing as thinking. *Language sciences*, 29(5), 621-632.
- Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences*.

Oostdam, R. (2005). Assessment of Argumentative Writing. In: Rijlaarsdam, G., van den Bergh, H., Couzijn, M. (eds) *Effective Learning and Teaching of Writing. Studies In Writing*, vol 14. Springer, Dordrecht. [https://doi.org/10.1007/978-1-4020-2739-0\\_29](https://doi.org/10.1007/978-1-4020-2739-0_29)

OpenAI. (2023). ChatGPT (Version 3.5) [Large language model]. <https://chat.openai.com/chat>

OpenAI. (2024). ChatGPT (Version 4o) [Large language model]. <https://chat.openai.com/chat>

Preiss, D. D., Castillo, J. C., Grigorenko, E. L., & Manzi, J. (2013). Argumentative writing and academic achievement: A longitudinal study. *Learning and Individual Differences*, 28, 204-211.