

Reviving the Philosophical Dialogue with Large Language Models

ROBERT SMITHSON
ADAM ZWEBER
UNC Wilmington

Abstract: Many philosophers have argued that large language models (LLMs) subvert the traditional undergraduate philosophy paper. For the enthusiastic, LLMs merely subvert the traditional idea that students ought to write philosophy papers “entirely on their own.” For the more pessimistic, LLMs merely facilitate plagiarism. We believe that these controversies neglect a more basic crisis. We argue that, because one can, with minimal philosophical effort, use LLMs to produce outputs that at least “look like” good papers, many students will complete paper assignments in a way that fails to develop their philosophical abilities. We argue that this problem exists even if students can produce better papers with AI and even if instructors can detect AI-generated content with decent reliability. But LLMs also create a pedagogical opportunity. We propose that instructors shift the emphasis of their assignments from philosophy papers to LLM dialogues: philosophical conversations between the student and an LLM. We describe our experience with using these types of assignments over the past several semesters. We argue that, far from undermining quality philosophical instruction, LLMs allow us to teach philosophy more effectively than was possible before.

Key words: large language models, artificial intelligence, philosophy papers, LLMs, dialogue

1. In Praise of the Undergraduate Philosophy Paper

The undergraduate philosophy paper is a beloved genre of assignment taking many forms. In many advanced courses, students write a traditional term paper: a work that resembles, at least in a rudimentary way, formal academic writing. In many introductory courses, instructors assign shorter papers with established topics and structures. Consider, for example, the classic “argument/objection/response” prompt:

Present the causal exclusion argument against dualism. Provide what you consider to be the strongest dualist response to this objection. Provide a response on behalf of the physicalist.

These assignments are venerated for good reason: they help students practice many important philosophical skills. For example, to make their papers easy to follow, students must practice organizing their thoughts. To prevent misunderstandings, students must practice communicating clearly and precisely. To make their position persuasive, students must practice anticipating and responding to objections. Perhaps best of all, philosophy papers help students to track a sustained argument in the direction of greater depth.

The ability (and desire) to follow a sustained argumentative thread is a mark of serious philosophical engagement. It is the mark of someone who wants to move past appearances and get to the heart of the matter. It is the mark of someone who wants “to get to the bottom of things” and who wants “to put everything in its proper place.”

It is also a difficult skill. Think of how Socrates must hold his interlocutors to account. They would jump quickly from one argument to the next, but Socrates would have them slow down and see a line of reasoning all the way through. When encountering an obstacle, they would rather change subjects or even walk away, but Socrates implores them to follow the thread a bit further.

This is one reason why philosophy papers, which ask students to track a sustained argument, are wonderful assignments. A good philosophy paper does not merely report a set of facts; nor does it merely list the considerations for and against a given position. Rather, it provides a structure revealing how the arguments in a certain area “hang together.” We tell students that, in these papers, they are not supposed to mention every idea that has occurred to them. Instead, they should take up a single thread of argument, tracing that strand without being distracted by others.

For example, in the above prompt, we do not ask students to mention every dualist response covered in class. Instead, we ask them to develop one such response so that they can explain it more carefully. In the final part of the prompt, we do not ask students to provide some entirely independent objection to dualism. Instead, we ask that this section directly respond to the discussion coming before. We remind students to emphasize depth over breadth.

At first, most students find it difficult to write good papers. But, by working through these challenges---by working out the best way to articulate objections, by struggling for the clearest way to frame a disagreement, by molding an inchoate idea into words---students practice and develop a variety of important intellectual and philosophical skills.

2. *The Crisis*

This, at least, was the hope. But many worry that large language models (LLMs) subvert the undergraduate philosophy paper, at least as traditionally conceived. One might view this development with enthusiasm or dismay.

For the enthusiastic, LLMs merely subvert the idea that students ought to write philosophy papers entirely on their own. On this view, students should be encouraged to use LLMs when writing papers: LLMs will help students produce better work.

On a more pessimistic stance, LLMs merely facilitate plagiarism. In response, some instructors have considered new types of “AI-proof” assignments.¹ Others advocate for returning to in-class assessments.²

We believe that the above controversies neglect a more basic crisis. This crisis exists even if LLMs can help students write better papers; it exists even if instructors can detect AI-generated content with decent reliability. The crisis is as follows:

1. Students can, with minimal philosophical effort (i.e., effort that develops the philosophical abilities mentioned in section 1), use LLMs to produce papers that are at least superficially impressive.
2. If (1) is true, then many students will use LLMs in this minimal-effort way (even if instructors take steps to discourage such use).
3. Producing a paper with minimal effort does not develop a student’s philosophical abilities.
4. Crisis: many students will, when using LLMs, fail to develop their philosophical abilities when producing undergraduate papers.

We comment on each premise in turn.

2.1. Premise 1

It is difficult to make any general claims about the quality of AI-generated philosophy papers. Fortunately, we can set aside every question about how good or bad AI-generated philosophy papers are (either now or in the future). For the sake of argument, let us simply assume that LLMs write poor philosophy papers and that future LLMs will do no better. Even so, it remains true that, with minimal effort, one can use LLMs to produce outputs that are at least *superficially* impressive. By this we mean: LLMs can produce outputs that seem impressive to undergraduates who have never taken a philosophy class before.

How little effort is required? The simplest method is simply to feed the LLM a verbatim copy of the paper prompt: a few seconds of mindless activity. Savvier students will run the resulting output through a paraphrase program that helps mask the text's artificial origin (a few further seconds of mindless activity). Alternatively, students can paraphrase the output on their own, sentence by sentence.³ While this method requires some effort, it still requires minimal *philosophical* effort.

2.2. Premise 2

If premise (1) is true, then many students will be tempted to use LLMs in a “mindless” way. Will they do so?

Many students will not. They properly appreciate the value of putting serious, disciplined effort into academic work. We would especially hope that this holds for students taking upper-division philosophy courses: students who have already experienced the value of struggling through writing philosophy.

But, many students, and especially students in introductory courses, may not yet appreciate this point. Some students have never encountered academic philosophy before and thus do not understand its value. Some students take philosophy just because it meets a general education requirement. Some students view college in general as a “hoop to jump through”: a mere credentialing process. And some students have no clear idea of why they are taking philosophy at all. These students are now asked to write a special kind of paper that is unlike anything they have written in the past and unlike anything they are likely to write in the future. What is the point of this?

Good philosophy courses will show students the value of writing philosophy. They may even show students the value of learning in general. But this process takes time and practicing writing is itself part of the process. So, at least initially, new philosophy students may need other incentives to struggle through the obstacles. Compare: a novice may not appreciate the value of practicing piano. Hopefully, the spark is kindled. But, at first, they may need structure to keep them practicing.

Accordingly, perhaps there are steps that instructors can take to discourage the mindless use of LLMs. Here are some steps that we took in recent courses:

1. *Curated assignments*: We used assignments that are difficult for current LLMs.⁴
2. *Discouraging external sources*: We discouraged students from using external sources; in fact, we prohibited students from using external sources that were not approved beforehand. We thought this would discourage the use of

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

LLMs, which, without careful guidance, often draw on material not covered in class.

3. *Explaining the shortcomings of LLM content:* During class, we gave students' actual essay prompts to chatGPT. We then explained why, despite their impressive appearance, these AI-generated papers failed to do what philosophy papers ought to do: track a sustained argumentative thread. In this way, we hoped to clarify the difference between a good philosophy paper and the mere appearance of one.
4. *Warnings:* We clearly stated, in class and on the syllabus, that AI-generated papers constitute academic misconduct. We also told our classes about the many academic infractions involving LLMs that we had already prosecuted.

Despite these measures, many students in our recent courses still submitted (minimal effort) LLM papers. How could this be? For example, how would students use LLMs even after seeing that current LLMs struggle with their actual course assignments? In hindsight, this behavior is not surprising:

- (a) Students may get conflicting messages over when it is appropriate to use LLMs.
- (b) Students may have success submitting AI-generated work in other courses.
- (c) Despite explanations of their shortcomings, LMM papers may still *look* impressive to students. They probably look more impressive than anything the student could produce via traditional plagiarism.
- (d) Despite warnings, students may still believe that LLM papers are less risky than traditional plagiarism.
- (e) Perhaps most crucially, it takes even less effort to generate a philosophy paper with LLMs than it does to engage in traditional plagiarism.

On this basis, we expect that, *even if instructors take steps like (1)-(4) to discourage the use of LLMs*, the incidence of AI-generated content will be higher than the incidence of traditional plagiarism (both historically and in the future). This aligns with our observations from recent classes, although this is a small sample size. We expect that there will be more data on this question in coming years.

The issue of detecting AI-generated content deserves special comment. There has been disagreement over how difficult it is, and will be, to detect such content. But we think that this controversy neglects a more fundamental issue. For the sake of argument, let us assume that

it is (and always will be) reasonably easy to detect the use of LLMs. *It does not matter*: many students will still submit minimal effort LLM papers. While this is an empirical matter, we base our prediction on considerations like (a)-(e) described above.

2.3. Premise 3 and the Crisis

Premise 3 requires little defense. Suppose that a student copies a course prompt verbatim into an LLM to generate their paper. Clearly, this does nothing to improve their philosophical abilities. Suppose that a student generates a paper with an LLM and then paraphrases the result (see section 2.1). This also does little to improve their philosophical abilities. Similar remarks apply to other low effort methods that students might use to generate an LLM paper.

Hence, we have a crisis. LLMs enable students to produce superficially impressive philosophy papers with minimal effort. Many students will use LLMs in this way (even if instructors take steps to discourage such use). These students receive little educational benefit as a result.

As mentioned earlier, this crisis is independent of two other controversies.

Controversy 1: Can LLMs help students produce better philosophy papers?

For the sake of argument, let's assume that they can. Even so, there is a crisis. This is because the main value of philosophical writing assignments is not the "product." (We do not ask introductory students to write papers on utilitarianism because we think that they may develop some novel argument for or against this view.) Instead, they are valuable because they help students practice certain cognitive skills.

Indeed, the crisis exists even if, in ideal cases, LLMs *help* students develop philosophical skills when writing papers. For example, imagine the best possible case: a student uses an LLM to help anticipate objections to their thesis, they carefully reason through the LLM's responses, and so on. This student, we imagine, will become a better philosopher through their effort. Even so, it remains true that students can produce superficially impressive papers without such philosophical effort. So, while LLMs may help some students, they only impede others.

Controversy 2: Can we reliably detect AI-generated content?

From the standpoint of fairness, this is an important question. We would not want a student who merely copies a prompt into an LLM to receive a better grade than someone who struggled diligently through a paper on their own. But let's suppose, for the sake of argument, that we can reliably detect AI-generated content. Then we may assign the

right grades. Still, a crisis remains: many students will not benefit intellectually from their assigned papers.

3. The Root of the Crisis

We have a crisis. But the crisis is nothing new. LLMs have merely exposed cracks in the undergraduate philosophy paper that existed long before.⁵ To see this, consider the following variant on the argument from section 2 that does not mention LLMs at all:

- 1* Students can, with low philosophical effort, produce papers that are at least superficially impressive.
- 2* If (1*) is true, then many students will produce papers with low philosophical effort (even if instructors take steps to discourage such papers).
- 3* Producing a paper with low philosophical effort does not significantly develop a student's philosophical abilities.
- 4* Crisis: many students will fail to significantly develop their philosophical abilities with philosophy papers.

This variant is itself plausible. To support (1*), the reader might suppose that we are thinking about traditional plagiarism. But, in fact, we think that (1*) is true even when considering students who write papers without any misconduct. To show this, consider the following cases:

- A. Student A gives ChatGPT a paper prompt and turns in the result.
- B. Student B writes a paper that closely mirrors the structure of a sample paper provided by the instructor. She substitutes in her own examples and makes other similar adaptations.
- C. Student C writes about the first objection that comes to their mind. It turns out to be an important issue. They also consider an important but obvious rebuttal. Again, it was the first and only reply that came to mind.
- D. Student D tries to mention *every* idea that occurs to him.
- E. Student E, thinking that the paper should resemble research papers in her other classes, finds a difficult article about the paper topic online. She stumbles through the exposition of this article, relying on many long quotations.
- F. Student F's paper is essentially a list of all of the points mentioned in the lecture notes.

- G. Student G, fancying herself a master at “bullshitting,” dashes off a paper filled with jargon and attempts to flatter her professor’s political sensibilities.

Just as in section 2.1, we can set aside all questions of how good or bad the above papers are. Some of them, considered purely as “products,” may be good (see, e.g., students B and C). Others are poor. Regardless, we can imagine that all of these papers at least *seem* impressive to the students writing them. In other words, we can imagine that all of the above students think that they are turning in “what the instructor is looking for.”

Crucially, none of the above papers required significant philosophical effort: effort at the kinds of philosophical skills mentioned in section 1. Some required no effort of any kind. Some required effort, but not effort of the right kind.

As for (2*), we trust that this premise is confirmed by the reader’s own teaching experience. Below in section 3.1, we discuss why so many students submit papers that do not involve the proper kind of philosophical effort (despite the student *thinking* that they are submitting impressive work).

Premise 3* is much the same as Premise 3 from the earlier argument. In the above list, some students will benefit from the assignment more than others. But, because none of these students put in the right type of effort, the paper was not an effective assignment for any of them.

3.1. A General Problem with the Philosophy Paper

The above argument reveals a general problem with the undergraduate philosophy paper. The problem is not just that students can use LLMs to produce such a paper. Nor is it even a problem with academic misconduct: many of the above students are not guilty of any such thing. The problem is not even about student effort *per se*; we can imagine, for example, that students D and E work very hard indeed. The problem is more general: students often fail to properly understand the value of philosophy papers.

This is true in each of cases A through G. Some students do not see philosophy papers as valuable at all. These students will be tempted to take the path of least resistance. Perhaps this now involves LLMs. But even if students do not use LLMs (because of their integrity, or because of clever assignments, or . . .), they will still be tempted to write papers like Student B (who unthinkingly copies a template), or Student C (who dashes off whatever first comes to mind), or student F (who tries to flatter her professor’s political sensibilities).

Other students will trust that philosophical papers are valuable, but fail to understand what that value is. (These students may produce

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

a research paper, or a mere summary of lecture notes, or . . .) And still other students will understand the abstract point that philosophy papers develop one's cognitive skills, but will still fail to understand how papers accomplish this. (These students may produce a mere list of objections, or . . .) All of these students may work hard, but their effort is misplaced.

Of course, instructors try to prevent this. We spend time in class discussing resources for philosophical writing. We discuss sample papers from previous classes. We explain, with sincerity and passion, the importance of developing philosophical skills. But, even so, it is evident that many students fail to properly understand the value of philosophy papers.

We now argue that this general problem results from two fundamental flaws with paper assignments.

3.2. The First Flaw: Idiosyncrasy

The first main flaw is the philosophy paper's *idiosyncrasy*. Looking towards the future, students know that they are unlikely to write anything like a philosophy paper again. Thus, it is natural for students to ask why these papers are so important.

Looking towards the past, students come to our classes with years of writing habits drilled into them from all kinds of sources. Now they are told that philosophy papers are different from the papers they have written before, and not just because they can use first-person pronouns. Still, it is a tall order for a single class to change their conception of what a paper aims to do.

Moreover, there are opportunity costs to devoting instructional time to explaining what these idiosyncratic assignments are supposed to look like. We would hope that, at least for introductory classes, we could use assignments that do not require a week's worth of class time simply to explain how to complete them.

3.3. The Second Flaw: The Gulf between Product and Process

The second, and deeper, flaw of the undergraduate paper is the gulf between the ultimate product and the thought processes underlying it. We will call this the "thought-process gulf." We can see this gulf by again considering students A through G. All of these students, we imagine, submit papers that they think are impressive. But the underlying thought process is very different in each case. Indeed, we can even imagine that some of these students, such as students A through C, submit very similar papers despite very different underlying thought processes. There are two ways in which this gulf is problematic.

First, this gulf is another part of the explanation for why students often fail to properly understand the value of philosophy papers. If students could directly see the proper thought process underlying philosophy papers, then many would understand the goal of these papers. They would probably even understand why this goal is important. But what students directly see is a product. For some students, this product will seem to resemble products they submit for other classes. These students may mistake the philosophy paper for a research paper, or a personal reflection piece, or something else familiar to them. For others, this product will seem to be governed by its own opaque conventions. When we explain these conventions (“Define your terms!,” “Write ‘In this paper, I will argue that ____,’” etc.), students may mistakenly judge that our main concern is just the product itself. Some of these students will look for the “trick” (e.g., follow the formula of the sample paper, copy everything from the lecture notes). Others will wonder why they should care about this peculiar product at all.

Even students who value philosophy may question this. We are all familiar with students who make excellent contributions to class discussions, but who struggle to grasp the conventions that govern papers. These students may wonder: why are papers so different from the (exciting, enjoyable, important) activity from class? Why follow these stolid conventions that poorly reflect my thought?

Second, this gulf is the very reason why it is possible for students to submit papers with the wrong kind of effort in the first place. Consider student H:

- H. Student H writes a paper in several drafts. They try out several different structures, thinking carefully about which is clearest. They think through several objections, but realize that some of them are not persuasive. Ultimately, they think of an impressive objection and include it in their paper. They consider several decent replies but, being an intro student, are unsure of their relevance. Ultimately, they focus on the most obvious (yet important) reply.

Student H writes a good paper because they put in serious philosophical effort. We want students to write their papers like student H. But, because of the thought process gulf, it is possible for students like A through G to submit papers that at least “look good” without the right kind of effort. Indeed, students A through C might, in principle, submit the same paper as H without this effort.

Thus far, we have viewed the thought process gulf from the student’s perspective. But, from the instructor’s perspective, this gulf manifests as a diagnostic problem. For example, when reading a paper, we may be unsure whether a student really understands a certain argument. We

would like to ask further questions to test this understanding, but the text cannot answer. Indeed, in the *Phaedrus*, Plato himself laments this feature of philosophical writing: the fact that written words cannot reply to questioning. For Plato, written philosophy was always second best.

We have described certain problems with the thought-process gulf. But, before moving on, we should acknowledge that this gulf also has certain advantages. For example, the thought process underlying a paper can be messy and unruly. But, through editing and revising, we can create a product that is more clear and more readable. In addition, the editing process can itself be an effective method for working out and refining our own ideas. Even so, as instructors, we often wish that we had a better window into how a student is engaging with a philosophical topic.

4. *In Praise of the Philosophical Dialogue*

If written philosophy was always second best, what philosophy is better? For Plato, the best philosophy is the kind exemplified in active, critically-engaged dialogue. Says Murdoch (1992: 404):

Some might say that philosophy is certain arguments in certain books, but for Plato . . . philosophy is essentially talk. *Viva voce* philosophical discussion . . . is the purest human activity and the best vehicle of truth. Plato *wrote* with misgivings because he knew that truth must live in present consciousness and cannot live anywhere else. . . . A book cannot answer back or distinguish wise or foolish readers. It needs its parents to speak for it.

Most philosophers will agree with Plato on the value of talking through philosophy. This is why we concentrate on encouraging discussion in class. This is why the Q&A is so prominent in philosophy colloquia.

There are many advantages to talking through philosophy, especially when compared to the papers of section 3. A paper is liable to misinterpretation, but an interlocutor can address misunderstandings as they arise. A paper may not anticipate an objection, but an interlocutor can react to new and unforeseen critiques. A paper (once submitted) cannot change its mind, but an interlocutor can adjust her thinking towards greater nuance. A paper is the same for anyone who reads it, but an interlocutor can tailor her remarks so that she best engages with the specific person with whom she speaks.

There is an extensive literature on the pedagogical virtues of dialogues.⁶ But, most relevant to our purposes, dialogues do not share the two main flaws of undergraduate philosophy papers (see sections 3.2 and 3.3). In their place, dialogues have two special virtues.

4.1. The First Virtue: The Social Character of Philosophy

Philosophical dialogue is not at all idiosyncratic. For one thing, students in well-taught philosophy classes (of appropriate size) will witness quality philosophical dialogue in every class meeting. But, more importantly, dialogues are continuous with student experiences outside the classroom. Whether in the dining hall or across the dining room table, it is common for students to talk through philosophical issues with friends and family. In short, unlike philosophy papers, most students are already familiar with philosophical dialogue, even if not by name.⁷

In place of the flaw of idiosyncrasy, dialogues have a virtue: they manifest the social character of philosophy. The aim of philosophy is not, or is not only, some rarified state of private contemplation. Philosophy seeks after wisdom, but whatever wisdom it finds, it seeks to share. We hope that our students will become wise; we also hope that they will help others become wise. Supposing that our students acquire some measure of philosophical wisdom, how will they, in the future, exercise and share that wisdom?

For most students, it will not be through writing papers or writing of any kind. It will be through talk. Some years from now, one of our former students will have a friend going through a difficult divorce. The two will talk. They try to “get to the bottom of things,” considering ways of describing and explaining what transpired. With careful steps, the student helps her friend understand herself and her situation more clearly. The friend comes away less lost and hopeless.

Another student will serve on a jury in a delicate case. She will talk with her fellow jurors; she will explain subtle distinctions, ask careful questions, and call the group back to the crux of the issue. Another student will have a daughter who asks difficult and honest questions. The student will talk with her daughter: she will play devil’s advocate and ask difficult and honest questions in return. Certainly, writing philosophy papers could help a student prepare for these conversations. But perhaps the better preparation is philosophical conversation itself.

4.2. The Second Virtue: Proximity to the Active Mind

The second virtue is one emphasized by Murdoch and Plato: with dialogue, we get as near as we can to a thinker’s active mind, which is where true understanding lives. With written words, there is a gulf: a clear paper can veil a disordered mind and a clear mind can produce a disordered paper (see section 3.3). But, in dialogue, someone’s thoughts can be scrutinized and tested. We can examine not just what they do say, but what they would say. By prompting an interlocutor with new ideas and considerations, we “see” their thought process active before us. (*Viva voce*: “with the living voice.”)

With papers, the thought-process gulf makes it possible for someone to think that they are doing a good job with a paper without exerting the right kind of effort (see section 3.3). By contrast, it is more difficult to imagine this occurring during a philosophical dialogue with an experienced interlocutor. This is for two reasons. First, students already have a good sense of how dialogues are supposed to work because dialogues are already familiar to them (see section 4.1). Second, with dialogues, an experienced interlocutor can immediately point out misplaced efforts. A long list of objections can be met with a request to focus on just one. A summary of lecture notes can be met with questions that go beyond class discussion. Bullshit can be immediately called out. (Of course, a student may continue bullshitting. But, if the interlocutor keeps calling this out, it will be readily apparent to everyone that the student is failing to meet the expectations for the dialogue.) In this way, dialogues seem designed to safeguard against the kinds of misplaced efforts so commonly seen in papers.⁸

This is also why dialogues are more useful than papers when giving students diagnostic feedback. If we cannot tell whether a student properly understands a point after reading a draft of their paper, we might invite them to office hours where we can better assess the student's thought process.

5. The Difficulties of Dialogue

We all know the value of philosophical conversation. Indeed, for many of us, encouraging conversation is the most important element of our teaching. Why, then, do our assessments take such a different character?⁹ The short answer is that it is simply too difficult to assess students on conversation in a fair, helpful, and practical way. This can be seen by considering several forms such assessment might take.

5.1. Discussion in Class

To be sure, many classes include a discussion/participation element. But this is usually a small component of the student's grade. In online courses or larger sections, using dialogue as the primary assessment is not even an option. But even for small sections, there are intractable difficulties. First, there are concerns about fairness and equity. Some students are confident and charismatic speakers; for others, public speaking is very stressful. Some students are extroverted, likable, and funny; others may be shy, thorny, or aloof. In each case, assessing students on the basis of conversations may unfairly disadvantage the latter group.

Second, discussion in class is not comprehensive enough. We want students to practice tracking a sustained argumentative thread (see section 1). To do this, a given student must engage in a sustained

conversation. But, during class, any given student is not the primary interlocutor for long enough to provide this kind of practice. Nor is she the primary interlocutor for long enough for an instructor to provide her with detailed, specific, and nuanced feedback. (Compare the feedback that one can provide on papers vs. the feedback that one can provide on a student's contributions to the day's discussion.)

Nor would we want a single student to be the primary interlocutor for an extended period during class. In class, we want the students as a group to have a sustained conversation, not for the conversation to center on one student.

There are other tensions as well. We do not want students to feel the pressure of grades during class; we want them to feel free to experiment with new ideas. (We also want this to be an enjoyable, liberating experience for them.) In addition, we as instructors do not want to focus on grading during class. Our attention is better focused elsewhere: listening carefully to what students are asking, thinking of the clearest way to explain ideas, helping to get more students involved, etc.

5.2. Student-Instructor Conversations (i.e., Oral Exams)

Oral exams would provide a more comprehensive sense of a student's philosophical thinking. But there are problems here as well. Fairness and equity remain concerns. In addition, there is simply not enough time to schedule several one-on-one meetings with every student. This would be easy for a small graduate seminar, but it is not possible for someone teaching three or four courses with 30+ students per class.¹⁰

Finally, there are concerns about imperfect signals. A live conversation tells us about a student's thought at the moment, but it may not reveal much about their thought in general. On any given day, students might feel ill, be distracted by other exams, or uncharacteristically "blank." In all such cases, one would rather not assess students on the basis of a brief, high-pressure performance. (Philosophical ability involves thinking slowly and carefully; it is different from a quick wit.) Here is one place where philosophy papers are superior as assignments: students can write a paper as carefully as they wish, submitting only their best and most refined work.

5.3. Student-Student Conversations

On a third option, students have dialogues with another student and then submit a recording of it as their assignment.¹¹ This is an appealing idea: many students *already* have late-night dorm room conversations about philosophical topics (see section 4.1). With the option to re-record dialogues, students would have additional control over putting their best efforts forward. In addition, this option is potentially more practical than oral exams. While it would be impossible to listen to one

hundred audio recordings per assignment, students could use software to generate a transcript. (They might also be asked to edit this transcript for relevance and clarity.)

However, a new difficulty arises. In class, or in an oral exam, an instructor can step in (if necessary) to gently guide the discussion. The instructor can also point out salient issues that the student may not have considered, to which the student can then respond. Without this guidance, there is a substantial risk that many student-student dialogues will go off track. These dialogues would neither exhibit a student's abilities nor allow an instructor to provide the most useful feedback.

6. Reviving the Dialogue with LLMs

But, here, we think that there is an opportunity. While LLMs have exacerbated underlying flaws with the undergraduate philosophy paper (see sections 2 and 3), we think that they can revive the potential of the philosophical dialogue as a pedagogical instrument.¹² We propose that, at least in introductory classes, instructors shift the emphasis of their assignments from philosophy papers to LLM dialogues: philosophical conversations between the student and an LLM.

6.1. The Promise of LLM Dialogues

We have both used this type of assignment in our recent courses. (Collectively, we graded roughly 500 LLM dialogues in 2023.) For example, one of us asked students to have an ethical discussion with chatGPT that starts with the following prompt:

Ellen is 85 years old and is dying from leukemia. Although she is still lucid and in control of her faculties, Ellen is expected to live only a few more days. She is visited in the hospital by her daughter Denise and Denise's two children. Denise has asked her friend Susan to bring food over to the hospital for the family.

Susan and her daughter arrive with the food, and the two stay to chat for a while. When the conversation turns political, Denise becomes nervous. This is because she is afraid that her mother will say something offensive. While Ellen is generally kind to the people around her, she grew up in a prejudiced household. For this reason, there have been many times in the past where Ellen has made offensive remarks about members of other ethnic and religious groups. Denise has repeatedly confronted her mother about these attitudes in the past, but it has never seemed to help. Indeed, such confrontations have typically led to bitter conflict and hateful words.

Denise's worst fears are realized; in the course of a rant, her mother makes several offensive statements about Jewish people. To make matters worse, Ellen is unaware that Susan's family is Jewish. Susan and her daughter listen in

ROBERT SMITHSON AND ADAM ZWEBER

stunned silence. Denise is torn over whether she should confront her mother or try to quickly change the conversation.

—

Let's have an ethical dialogue about the above case. In this dialogue, you take the position that Denise should *not* confront her mother. I will take the opposite position.

Let's start by you presenting what you think is the single strongest consideration in favor of *not* confronting the mother.

To show the promise of this assignment, here is an excerpt¹³ of an ensuing LLM dialogue submitted by a recent student (chatGPT speaks first):

ChatGPT: Preserving peace and harmony: Confronting Ellen about her offensive remarks may disrupt the peaceful and harmonious environment necessary for Ellen's comfort during her final days. Engaging in a confrontation could escalate tensions, leading to unnecessary conflict and emotional distress for both Ellen and Denise. In this delicate time, it might be more beneficial to focus on maintaining a sense of calm and tranquility. . . .

Student: These are good arguments.

While a relationship with Denise's mother is important due to her final moments, a relationship with Susan will be longer than the time Denise has left for her relationship with her mother. Ultimately, Denise's relationship with Susan could provide her with more happiness than the relationship with her mother due to the length of time. Not confronting her mother and standing up against discrimination could potentially harm the relationship between Denise and Susan.

Another point to consider would be Susan's daughter in the room. Because Susan's daughter is from the younger generation, she is still observing society and how they react to people who are different. Susan's daughter might learn to be ashamed or embarrassed by her family heritage and religion if Denise's mother is allowed to continue to make offensive remarks without any confrontation. . . .

ChatGPT: Relationship longevity: While the relationship with Susan may potentially last longer than the remaining time with Denise's mother, it is crucial to recognize that relationships are complex and multifaceted. Denise's relationship with her mother encompasses a lifetime of shared experiences, love, and familial bonds. Prioritizing the current moment and focusing on maintaining a

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

peaceful environment during Ellen’s final days can help Denise find closure and ensure that their relationship ends on a positive note, regardless of its duration. . . .

Student: These points are all valid to consider.

While it is important to provide an positive note for their relationship to end on, would it truly be positive if Denise is not honest? Ellen’s remarks clearly made Denise uncomfortable and upset enough in the past to make a comment and attempt to change Ellen’s views. This occurrence could be no different and make Denise uncomfortable. If Denise does not speak up and pretends that everything is all right and she is happy with her mother, then are the final moments of their relationship truly positive due to the lack of honesty and genuineness being displayed by Denise.

ChatGPT: Emotional well-being and closure: Denise’s decision not to confront her mother in this specific instance does not necessarily imply a lack of honesty or genuineness. It recognizes the complexity of the situation and the emotional well-being of both Denise and her mother during Ellen’s final days. Denise may choose to prioritize finding closure and maintaining a sense of peace in their relationship, rather than engaging in a potentially contentious conversation that may not yield positive results. The positivity in their final moments can be derived from fostering love, support, and understanding, rather than focusing solely on challenging deep-rooted beliefs.

We offer several observations:

First, and most importantly, the above student is successfully practicing philosophy. In particular, they are practicing the crucial skill of tracking an argument in the direction of greater depth (see section 1). The conversation starts with a straightforward observation about Ellen’s happiness in her final days. But, after several exchanges, the student and LLM are discussing deeper questions about what is valuable in friendships. (Is it the expected future together? Is it a past of shared experiences? Is it honesty? Honesty about what kinds of things?) The student has not “gotten to the bottom” of this issue, but they have had good practice from trying.¹⁴

Second, there is nothing idiosyncratic about the above dialogue (see section 4.1). It is transparently similar to students’ ordinary conversations. We doubt that this student (or any other student) had any difficulty understanding the purpose and value of this assignment.¹⁵

Third, the transcript provides a clear glimpse of the student’s thought process. Unlike with papers, there is no major thought-process gulf

(see section 3.2): the student effectively *submitted* their thought process (or something close to it) as their assignment. This shows how, with LLM dialogues, it is difficult for students to make a sincere effort at completing the assignment properly without practicing the relevant cognitive skills. (Of course, there is a worry in the background: could students simply use LLMs to complete both sides of the dialogues? We discuss this issue in section 6.5.)

To be clear, we do not say that students completing LLM dialogues will never make mistakes. Of course they will: they are practicing difficult skills. For example, in the above excerpt, it would be better to fully discuss the relationship issue before raising the new consideration about the presence of Susan's daughter (cf. note 13). Nonetheless, this student was clearly exerting the right type of philosophical effort when completing the assignment.

What we have yet to see is the analogue of the all-too-familiar paper that confuses the assignment for something that it is not. For example, we have not received dialogues that merely copy the lecture notes, or list arguments with no analysis, or pretend to be research papers. In other words, we have not yet had students like A through G: students who submit assignments that "look good" without exerting significant philosophical effort.

Of course, we acknowledge that it is *possible* for students to submit dialogues that, e.g., merely copy lecture notes, just as this is possible for papers. But there is a crucial difference between the cases. With papers, a student who merely copies lecture notes may genuinely think that they are completing the assignment well. This is because the goal of a philosophy paper is not always obvious to students (see section 3). But, as discussed in section 4, there are two reasons why dialogues are different. First, students already understand how dialogues are supposed to work. Second, the interlocutor's presence in the dialogue helps prevent misplaced efforts: students know that they must address the interlocutor's remarks (and not, e.g., just copy unrelated notes).

Again, we can still imagine students who merely copy lecture notes into a dialogue. (Perhaps, for example, they are rushing to finish the assignment.) But, for the above reasons, such students will almost certainly recognize that they are not completing the assignment correctly. Thus, they stand outside the scope of our argument. For papers, the crisis was that students can and will submit papers that at least *seem* impressive (to them) despite not practicing philosophical skills. We have argued that this same crisis does not extend to LLM dialogues.

To reinforce how LLM dialogues avoid the crisis, let us try to apply the argument from section 3 to LLM dialogues:

- 1†. Students can, with low philosophical effort, produce LLM dialogues that are at least superficially impressive.

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

- 2†. If (1†) is true, then many students will produce LLM dialogues in this low-effort way (*even if* instructors take steps to discourage such production.)
- 3†. Producing an LLM dialogue with low philosophical effort does not develop a student’s cognitive and philosophical abilities.
- 4†. Crisis: many students will fail to develop their cognitive and philosophical abilities when producing LLM dialogues.

In section 3.3, we explained how, in the case of papers, the thought-process gulf is the ultimate reason why it is possible for students to submit superficially impressive papers with the wrong kind of effort. In other words, the thought-process gulf ultimately explains why premise 1* is true for papers. But, because there is no similar thought process gulf for LLM dialogues, it is not similarly possible for students to submit superficially impressive LLM dialogues with the wrong kind of effort. So premise 1† fails.

There is one critical issue not yet discussed. We began this paper by discussing how LLMs exacerbate an underlying problem with philosophy papers. In effect, we have just now argued that there is no similar underlying problem with dialogues. But maybe LLMs still generate a crisis for dialogue assignments because students can use LLMs to mindlessly complete both sides of the dialogue. We discuss this issue in section 6.5.

6.2 The Assignment Itself

We hope that the reader now sees the pedagogical potential of LLM dialogues. But, thus far, have not described any specific details of this assignment. Here is one example: the assignment corresponding to the excerpt in Figure 2.

In this assignment, you will have ethical dialogue with either chatGPT or GPT-4 about the ethical case in Fig. 1. Please create an account on the OpenAI website. chatGPT is free. GPT-4 requires a \$20/month subscription. You can use either one, although it takes slightly more work to keep chatGPT “focused” (see “helpful hints” below).

Step 1: Setting Up the Dialogue

Give the AI the full case and then tell it which stance you would like it to defend. Ask it to present the single, strongest argument in favor of their position.

Step 2: Conducting the Dialogue

ROBERT SMITHSON AND ADAM ZWEBER

In the dialogue, you should raise objections to the AI's argument. You can ask the AI for clarification. You can point out considerations that the AI may have ignored. This assignment is not graded on who "wins the debate." Instead, we are focusing on the following skill:

Tracking a "sustained thread of argument": try to directly respond to whatever the AI just argued (rather than bringing up new issues). Each new exchange in the debate should connect with what came before (like in a normal conversation).

Step 3: Editing the Dialogue

Copy your dialogue into a document. Then cut it down to four pages using the following criteria:

- At least one third of the dialogue should be your contribution.
- Don't change the writing, just cut out sentences. When you do so, put in an ellipsis (" . . .")
- Try to edit the dialogue so that the final product best showcases the above skill.
- In your submission, include a link to the unedited transcript through the OpenAI interface.

Helpful hints:

- The AI can be longwinded. So tell it to keep its responses to 150 words or less. You may have to keep telling it this.
- The AI can have trouble focusing on one issue. So ask it to focus by saying things like: "Respond directly to the point just given; do not bring up a new consideration," "Provide a single objection to the argument I have just given."
- The AI can lose track of its position. If it does, remind the AI of what stance they are taking.

This is a new type of assignment and we recognize that it is far from perfect. Our aim is not to provide the definitive guide to LLM dialogues; we hope that instructors will experiment with this genre of assignment in order to refine and improve it. However, there are some useful features of the above assignment that deserve comment.

6.3. The Editing Process

Step 3 requires students to edit their dialogues. We find that this step makes dialogues easier to grade. (Indeed, we think that dialogues are less tedious to grade than traditional papers: there is more interesting variability.)

This step also has pedagogical benefits.¹⁶ As noted in section 3.3, the process of editing a paper can help a writer to refine their thinking. The same holds for editing dialogues. Students practice their philosophical skills "live" when chatting with the LLM. They then get

further practice when analyzing the dialogue as they edit it for submission, especially if they are asked to do so in a way that prioritizes clarity, depth, or whatever other virtue that the instructor would like to emphasize. (Think of a chess student who practices their tactics “live” in a tournament, but then analyzes the recorded game afterwards for further practice.)

Of course, the editing process reintroduces some distance between thought process and product for LLM dialogues. But we think that this proposal strikes the right balance: enough distance to prioritize clarity and intelligibility, but still enough proximity to reveal how well a student is engaging with the philosophical ideas. And, crucially, the gulf is still narrow enough to prevent the kinds of misplaced efforts that are often observed in student papers (see section 6.1).¹⁷

6.4. Reining in the Dialogue

As of December 2023, ChatGPT (in its standard GPT 3.5 version) suffers from two deficiencies. First, it is long-winded: it often responds to philosophical questions with responses of 500 words, making for stilted and one-sided dialogues. Second, it can be a poor philosopher. It often brings up many arguments instead of focusing on one. It often ignores its interlocutor’s objections and merely restates earlier points. It sometimes loses track of its position. These flaws can make LLM dialogues a mess, both for students and instructors.

We have addressed these issues in several ways. First, the editing process cleans up the dialogue. As discussed above, this makes grading easier while also providing students with further philosophical practice. Second, we encouraged students to upgrade to the GPT-4 version of ChatGPT, which we have found to be a much better and more concise philosopher. The downside, and the reason GPT-4 was not required, is its \$20/month subscription cost (as of December 2023). Nonetheless, we hope that, as LLMs improve, this issue will be further mitigated.

Finally, the assignment includes “helpful hints” for prompting ChatGPT. In our experience, adding the suggested instructions from Figure 3 helps create a more natural dialogue. Sometimes, chatGPT ignores these instructions. But, again, it is useful for students to practice recognizing when an interlocutor has “gone off track,” or “shifted the goalposts,” or “mixed up their position.” In effect, students get to act like a modern-day Socrates: someone who guides a discussion, keeping it from going astray.

6.5. Can LLMs Complete This Assignment?

Thus far, we have not discussed a critical issue: can an LLM complete the entire dialogue on its own? If so, students could use LLMs to

produce minimal effort dialogues and the assignment would face the same crisis as traditional papers after all.

To address this worry, we first offer a general report: we have seen a far lower rate of academic infractions with LLM dialogues than with papers. In 2023, we collectively graded roughly 500 LLM dialogues. There were only two cases where we even suspected an academic infraction. Only one student tried to use an LLM to complete the entire dialogue; the second case involved traditional plagiarism. In both cases, the infraction was painfully obvious.

By contrast, over the same period, we graded over 300 papers. Of these submissions, we successfully prosecuted dozens of infractions, most involving the prohibited use of LLMs. Many of these infractions were not obvious. In addition, there were further cases where we suspected, but were not certain, that students used LLMs.

This suggests that LLM dialogues are less susceptible to cheating than traditional papers. First, there are technical obstacles. At the very least, students cannot just ask chatGPT to produce a dialogue with a simple prompt, e.g., “Produce a philosophical dialogue between a student and chatGPT about X.” This is because the assignment requires students to provide a link to their original, unedited dialogue (see step 3 of Figure 3). OpenAI enables users to generate such a link; the link loads the dialogue *within the OpenAI interface itself*. (For an example, see note 4.) Thus, a student could not use the above prompt; the instructor would immediately see, upon clicking the link, that chatGPT generated the whole exchange.

Nor can a student recreate an LLM-produced dialogue simply by copying, step by step, one side of this dialogue into a new chat. LLMs typically produce unique outputs even when given the same prompt multiple times, and in this case, the prompts are different.

Thus, cheating would require some kind of incremental approach, e.g., ask chatGPT to begin a dialogue about topic X, copy this output into a new chat and ask chatGPT for a reply, copy this reply into the original chat and ask chatGPT for a rebuttal, copy this rebuttal into a new chat and ask for a new reply, and so on for every step of the dialogue.

Certainly, students could attempt to use such a method. But, as it happens, this is difficult to do in a convincing way. The reader can see this for themselves. Challenge: without using one’s philosophical abilities, use the above method to produce something that looks like a genuine philosophical discussion between an LLM and a student.

The difficulty is not merely stylistic. There are all sorts of “moves” that students naturally make when talking through philosophical issues:

- Requesting clarification of an idea or argument

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

- Calling out” the interlocutor for an argumentative misstep, such as “shifting the goalposts”
- Challenging the interlocutor to provide argumentative support
- Revising an argument to address an objection or misunderstanding
- Setting up an objection with a series of pointed questions

But, for one reason or another,¹⁸ it is not natural for chatGPT to make any of the above moves. (Again, readers can test this for themselves.) To be sure, one can *get* chatGPT to perform the above actions through detailed prompt engineering. But this requires more effort.

And, crucially, the effort in question is not mindless. Students cannot just copy and paste a generic request for chatGPT to call out a misstep (or to request clarification, or . . .); such instructions only make sense in appropriate contexts. But, plausibly, students must exercise their philosophical abilities in order to identify such contexts. (Hence we challenged the reader to use chatGPT to create an authentic-looking dialogue *without exercising one’s philosophical abilities*.) It is perhaps some consolation that, with LLM dialogues, convincing attempts to cheat still require the hard work of practicing philosophy. But, more importantly, this very fact makes it unlikely for students to cheat in this way. It is easier just to argue with chatGPT on one’s own.

It may be objected that the above remarks focus on contingent behaviors of current versions of chatGPT. We agree that there is no reason in principle why an LLM could not be trained to make the kinds of moves listed above. But we recommend crossing this bridge upon reaching it. There is no obvious reason why we should expect future LLMs to become more “naturally human” in their conduct during dialogue. Consider that the above behaviors—e.g., chatGPT’s reluctance to challenge its human users—may be regarded, at least by OpenAI, as features, rather than bugs, of the platform—see note 18.¹⁹

It is worth emphasizing that the dialogues naturally produced by chatGPT (i.e., dialogues constructed without sophisticated prompt engineering) look inauthentic not only to instructors, but also to students themselves. As discussed in section 4.1, students already know what it looks like to argue through issues. Accordingly, they can tell that chatGPT-produced dialogues are not “what the instructor is looking for” (which is, of course, a real conversation). Here, we again see a crucial difference between LLM dialogues and papers, which can be opaque to students (see section 3.2).

There are also positive incentives to complete the assignment honestly. As just mentioned, students already know how dialogues work. Indeed, they probably already understood why dialogues are valuable.

So dialogue assignments are less likely to seem like a silly “hoop to jump through” (see section 2.2).

In addition, many students told us how much fun they had arguing with chatGPT. (Far fewer students have told us how much they enjoyed papers.) This is not surprising: many people like to argue, at least when the dialogue is respectful. Beyond its role in discouraging cheating, the fun of LLM dialogues is an independent virtue of the assignment. Philosophy can be difficult, so anything instructors can do to make this activity enjoyable to students is to be commended.

In summary, we acknowledge that LLM dialogues, like perhaps any assignment, are not immune to cheating. But, when compared to papers, LLM dialogues are both more difficult to cheat and also more enjoyable to complete honestly.

6.6. Grading

Our purpose has been to show the promise of a new genre of assignment, rather than fully working out the specific details (about grading, etc.) of how best to implement LLM dialogues. We hope that instructors will experiment with the basic model of this assignment, tailoring it to fit the goals of their specific classes.

However, we here provide a template for a rubric, which we hope is useful to instructors who employ this type of grading scheme. The rubric emphasizes one set of possible learning objectives for LLM dialogues:²⁰

Table 1: Template for a Rubric

	Excellent	OK	Poor
Arguments	Student provides relevant responses to the claims made by the LLM. Students' objections are well-motivated and their reasoning genuinely supports their claims. Student emphasizes “depth” over “breadth.”	Student provides support for their views, but they sometimes fail to respond to the LLM's points. Student sometimes fails to raise appropriate objections to the LLM.	Student provides little or no support for their claims. Student does not respond to what the LLM actually says (by instead, e.g., raising independent issues throughout the dialogue).
Creativity	Student provides examples, questions, and arguments that are appropriate to the context and that go beyond what is covered in lecture or readings.	Student provides examples, questions, and arguments that are appropriate to the context, but these do not go beyond material covered in class.	Student fails to raise any independent line of thought, by, e.g. repeating points from the class even when such points are not relevant.

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

	Excellent	OK	Poor
Content Knowledge	When appropriate, student demonstrates understanding of course content by clearly explaining these issues in her dialogue. Student corrects confusion from the LLM. Student asks appropriate questions about topics that student does not recognize from class discussion.	There are minor mistakes relating to course content. Student makes little effort to clarify the LLM's remarks when something is difficult to understand.	There are major mistakes relating to course content. For example, student fails to recognize when chatGPT discusses an argument from class. Student makes no effort to clarify the LLM's remarks when something is difficult to understand.
Prompt engineering/control	<p>Student's prompts lead to LLM outputs that contribute to a productive conversation.</p> <p>When appropriate, student identifies missteps from the LLM (e.g., points where the LLM changes their position or merely repeats earlier points).</p> <p>Student works to get the LLM to "focus," i.e., to write concisely, to focus on a single issue at a time, and to reply to what the student has written (rather than bringing up unrelated points).</p>	<p>Student's prompts lead to LLM outputs that are not relevant to the discussion.</p> <p>In some cases, student fails to identify missteps from the LLM.</p> <p>Student makes little effort to constrain or focus the LLM's contributions to the dialogue.</p>	<p>Student fails to get the LLM to engage in philosophical dialogue, by (e.g.) merely requesting information.</p> <p>Student entirely fails to identify missteps from the LLM.</p> <p>Student makes no effort to constrain or focus the LLM's contributions to the dialogue.</p>
Editing	<p>The edited dialogue focuses on the most relevant and important parts of the exchange.</p> <p>Student edits the dialogue in such a way that the final product would be clear and intelligible to an unfamiliar reader. The edited dialogue explains technical terms and avoids distracting stylistic or grammatical issues.</p> <p>Student's side of the edited dialogue is roughly half of the final submission.</p>	<p>The edited dialogue does not always focus on the most relevant or important parts of the exchange.</p> <p>Student's edited dialogue would be difficult for an unfamiliar reader to understand. (While the instructor can follow, an unfamiliar reader could not.)</p> <p>Student's contribution is substantially less than half of the final submission.</p>	<p>The edited dialogue fails to focus on the most relevant or important parts of the exchange.</p> <p>The instructor is unable to understand the student's discussion (due to stylistic issues, grammatical issues or some other factor).</p> <p>Almost all of the edited dialogue was generated by the LLM.</p>

7. Conclusion

There are some philosophical skills that can only be practiced through the process of writing. When writing papers, one needs to be especially clear in one's language and argumentation because one is unable to immediately correct possible misunderstandings. Similarly, one needs to be especially skilled in anticipating objections because one's interlocutor is not present to mention issues that have been overlooked. With the proposed shift to dialogues, students may not receive the same practice in these areas.

For the above reasons, we do not advocate the abolition of the undergraduate philosophy paper. Instead, we recommend that instructors shift emphasis from papers to dialogues, *especially in introductory courses*. In intro courses, many students do not yet understand the abilities that are needed to do philosophy well, such as the ability to track a sustained argumentative thread or the ability to pursue an argument in the direction of greater depth. Nor have they yet learned to value these abilities. We have argued that LLM dialogues can help students practice these skills; indeed, we have argued that they are more effective at this than traditional papers.

Imagine that a student takes some introductory classes with LLM dialogues as assignments. If all goes well, this student comes to understand the virtues of good philosophy and the value of practicing it. Maybe, as a result, this student decides to take some upper-division philosophy courses; maybe they even decide to become a major. In those courses, they have the opportunity to write traditional papers (perhaps in addition to LLM dialogues).

To learn to write papers, the student will need guidance, just as before. However, we expect that the student's experience with LLM dialogues will make the process easier. This is because papers and dialogues share certain fundamental goals. The hope is that, because LLM dialogues have fewer opaque conventions, they provide a good first introduction to fundamental philosophical skills. Having grasped these skills, students can then more easily learn the conventions of paper writing without being distracted by them.

Indeed, we can imagine ways in which LLM dialogues could *directly* help students learn to write papers. For example, one might ask students to transform an LLM dialogue into a traditional paper. This would help clarify the fact that philosophy papers, like dialogues, aim to track a sustained argumentative thread. It might also help students with incorporating anticipated objections into their writing.

Of course, we have argued that LLMs raise a certain crisis for traditional papers. So, if instructors transition to papers, they may still have to consider strategies for deterring or detecting the mindless use of LLMs. We have little to add to the discussion of these issues.

REVIVING THE DIALOGUE WITH LARGE LANGUAGE MODELS

However, we would hope that experience with LLM dialogues in intro courses might make students less likely to cheat on traditional papers later on. As just discussed, LLM dialogues provide students with a good first introduction to fundamental philosophical skills. But, by practicing these skills, students will also (we hope) come to see why these skills are so valuable. They will then be more motivated to seriously practice these skills with traditional papers later on, as opposed to taking the low effort route. Here, we are optimistic because, in our experience, the mindless use of LLMs is far more common in introductory courses than in advanced courses for majors.

Notes

1. See, e.g., some of the suggested assignments in Stafford 2023.
2. See, e.g., Worthen 2022.
3. Students might do this to make the paper look less like the output of an LLM. See Terry 2023 for discussion.
4. See, e.g., <https://chat.openai.com/share/9dcfd5f3-6a65-4ca8-8893-0a5ae0245cb8>. This assignment invokes Hill's (1979) notion of a "purely symbolic protest": a protest that someone engages in, at significant personal cost, even though they recognize that their action will not cause any social change. We gave the May 24 version of chatGPT this prompt thirty times in different chats. In every case, its responses were irrelevant to Hill's discussion. For example, in the above transcript, chatGPT interprets a purely symbolic protest as one involving symbolic gestures.
5. Tellingly, Russell (2003: 77), describing the adoption of term papers by American universities in the late nineteenth century, writes "Because performance was measured primarily by final essay examinations, enterprising students produced and sold what Harvard faculty called 'unauthorized professional tutors' outlines': concise responses to questions or topics contained in the syllabi or old examinations of large lecture courses." The ability to produce low effort papers has been a problem since the invention of papers, not since the invention of LLMs.
6. For example, Lee (2018) discusses how dialogues can help students test the logical relations between their beliefs. L'Hôte (2012) argues that dialogues are particularly effective at persuading students of the value of philosophical thinking. Mullis (2009) discusses how dialogue can help encourage independent thinking. Coppenger (1979) specifically motivates dialogue assignments as a way to avoid a host of common problems with traditional papers. See also Baker, Andriessen, and Schwarz 2019.
7. This makes philosophical dialogues easier for students to understand than papers. But it also makes students more likely to value dialogue. It is a truism of pedagogy that students will take more interest in an activity or topic that they routinely encounter in their lives. They do so with dialogue: many students love to have these conversations. In contrast, typical philosophy papers are removed from both students' everyday lives and their academic past and future.

ROBERT SMITHSON AND ADAM ZWEBER

8. One caveat: as Walton (1989) and Baker and Schwarz (2017: chap. 3) note, there are different kinds of dialogue, some of which are difficult for beginners to distinguish. For example, a student might confuse a dialogue aimed at inquiry for one aimed at persuasion.

9. It was not always so. Russell (1991: chap. 3) notes that prior to the late nineteenth century, oral assignments were the primary form of evaluation in American universities.

10. Medeiros, for example, describes a clever “spoken exam.” But the described exam takes at least a week to administer, with sessions outrunning standard class time. He writes (2017: 41) that “For *Biomedical Ethics* with twenty-five students, one week of *The Spoken Exam* would require about seven hours; for *General Ethics* with one hundred students, one week would require between twenty-one and thirty hours.” Baker and Schwarz (2017: chap. 6) also note the practical difficulties in evaluating dialogues “live.”

11. Should the other student be a member of the class? If so, they will likely be better prepared for the dialogue. On the other hand, holding a conversation with an outside friend forces the student to find new ways of explaining her ideas and could get a new student interested in philosophy.

12. While the use of LLMs in dialogue is a novel idea, Baker, Andriessen, and Schwarz (2019) note that technology (such as chat rooms) has long played an important role in the use of dialogues in education.

13. In the full dialogue, the student and LLM discussed three separate issues (which they numbered 1–3) in each exchange. To be concise, our excerpt presents just one of these threads. We further comment on this issue below.

14. We invite the reader to consider how and whether the above ideas would have appeared in a traditional paper for an intro class.

15. Anecdotally, we have both received unsolicited positive feedback on LLM dialogues from students who found them valuable and also fun.

16. Howe, Hennessy, and Mercer (2019) note that a theme in educational research suggests that the adoption of a meta-cognitive perspective, in which students evaluate their own dialogues, promotes positive learning outcomes. Schwarz and Baker (2017: chap. 6) also note that computer-mediated dialogues, by providing the opportunity to re-read and revise student contributions, can promote deeper reflection on the part of students. We take both of these to be potential benefits of the editing step of the dialogue.

17. For reasons discussed in section 6.5, we also recommend that instructors require students to submit a link to their original unedited dialogue. This requirement makes the gulf even more narrow.

18. We speculate that some of this behavior is explained by the human feedback step of chatGPT’s training procedure. Suppose that, during this training step, human testers preferred “respectful” and “authoritative” outputs. This might explain why chatGPT is reluctant to “call out” their opponent in a dialogue, to request clarification of an idea, to challenge the interlocutor for more support, etc.

19. To be sure, someone could, in principle, train up an entirely new LLM, feeding it thousands of examples of actual student dialogues. Or they might just decide to argue with chatGPT on their own.

20. See Schwarz and Baker 2017: chap. 6 for a general discussion of how to evaluate dialogues. For a possible alternative set of learning objectives, see Van Eemeren and Grootendorst’s (1992) “10 Commandments” for dialogues.

Bibliography

- Baker, Michael, Jerry Andriessen, and Baruch Schwarz. 2019. "Collaborative Argumentation-Based Learning," in *The Routledge International Handbook of Research on Dialogic Education*, ed. Neil Mercer, Rupert Wegerif, and Louis Major, 76–88. New York: Routledge. <https://doi.org/10.4324/9780429441677-8>
- Coppenger, Mark. 1979. "Written Dialogue," *Teaching Philosophy* 3(2):197–202. <https://doi.org/10.5840/teachphil19793229>
- Hill, Thomas. 1979. "Symbolic Protest and Calculated Silence," *Philosophy and Public Affairs* 9(1): 83–102.
- Howe, Christine, Sara Hennessy, and Neil Mercer. 2019. "Classroom Dialogue and Student Attainment: Distinct Roles for Teacher-Led and Small-Group Interaction?," in *The Routledge International Handbook of Research on Dialogic Education*, ed. Neil Mercer, Rupert Wegerif, and Louis Major, 182–95. New York: Routledge. <https://doi.org/10.4324/9780429441677-16>
- L'Hôte, Crystal. 2012. "Philosophy in the Wild: An Introductory Exercise," *Teaching Philosophy* 35(3): 263–74. <https://doi.org/10.5840/teachphil201235328>
- Medeiros, Paul. 2017. "Introducing *The Spoken Exam!*," *Teaching Philosophy* 40(1): 37–55. <https://doi.org/10.5840/teachphil201733063>
- Mullis, Eric. 2009. "On Being a Socratic Philosophy Instructor," *Teaching Philosophy* 32(4): 345–59. <https://doi.org/10.5840/teachphil200932438>
- Murdoch, Iris. 1992. *Existentialists and Mystics*. New York: Penguin.
- Plato. 1997. "Phaedrus," in *Plato: Complete Works*, ed. J. Cooper, 506–56. Indianapolis: Hackett.
- Russell, David. 1991. *Writing in the Academic Disciplines: A Curricular History*, 2nd ed. Carbondale: Southern Illinois University Press.
- Schwarz, Baruch, and Michael Baker. 2017. *Dialogue, Argumentation, and Education*. Cambridge: Cambridge University Press.
- Staffel, Julia. 2023. "ChatGPT and Its Impact on Teaching Philosophy and Other Subjects," January 4. <https://www.youtube.com/watch?v=bkjVkfU9Gro>.
- Terry, Owen Kichizo. 2023. "I'm a Student. You Have No Idea How Much We're Using ChatGPT," *Chronicle of Higher Education*, May 2. <https://www.chronicle.com/article/im-a-student-you-have-no-idea-how-much-were-using-chatgpt>.
- Van Eemeren, Frans, and Rob Grootendorst. 1992. *Argumentation, Communication and Fallacies*. New York: Routledge.
- Walton, Douglas. 1989. "Dialogue Theory for Critical Thinking," *Argumentation* 3: 169–84. <https://doi.org/10.1007/BF00128147>
- Worthen, Molly. 2022. "If It Was Good Enough for Socrates, It's Good Enough for Sophomores," *New York Times*, December 2. Available at <https://www.nytimes.com/2022/12/02/opinion/college-oral-exam.html>.

Robert Smithson is an Associate Professor of philosophy at UNC Wilmington. He received his Ph.D. in philosophy from UNC Chapel Hill in 2016. His research interests include metaphysics, philosophy of mind, and the philosophy of science. smithsonr@uncw.edu.

Adam Zweber received his PhD in philosophy from Stanford in 2023 and currently teaches at UNC Wilmington. His research focuses on foundational issues in ethics, particularly metaethical naturalism. zwebera@uncw.edu.