Search is a Hammer, Generative Chat is a Loom; Beware the Technological Attribution Error

Bert Baumgaertner[1][0000−0001−8803−4771] and Zeth duBois[1][0000−0002−7505−069X]

University of Idaho, Moscow ID 83844 USA
bbaum@uidaho.edu, zdubois@uidaho.edu

Abstract. This paper explores the paradigm shift from traditional search engine usage to the emergent interaction with generative chatbots, specifically in the context of educational settings. We investigate how generative chatbots, exemplified by ChatGPT, represent a fundamentally different tool for inquiry and learning, resembling a loom for weaving thoughts, rather than a hammer for nailing down facts. Through a case study in an upper-division philosophy course on metaphysics, we analyze whether dialogues with ChatGPT can enhance the quality of argumentation and critical thinking among students. Our findings reveal that the conventional "one-shot" query approach, typical of search engine and encyclopedic interactions, significantly underutilizes the potential of chatbots. We argue for the need to reframe our interaction strategies with these AI tools, emphasizing the art of prompting and iterative refinement. The paper sheds light on the importance of adapting to the dialogical nature of generative chatbots as part of AI literacy and its implications for educational practices and cognitive development. Failure to do so may reflect a technological attribution bias.

Keywords: LLM · Chatbots · Generative AI · Pedagogy · Technological Attribution Error

1 Introduction

Generative chatbots like ChatGPT are quickly becoming complementary tools for both thinking and writing. By their very design they utilize an iterative method of inquiry, akin to a Socratic dialogue or the method of reflective equilibrium. The aim of this project was to assess whether the use of ChatGPT as a dialogical partner in a group assignment improves the quality of argumentation in an upper division philosophy course on metaphysics. The results are not what we expected, but the lessons we learned are important and generalize.

The mistake we made is, in hindsight, readily apparent. Our question emphasized features of a generative chatbot. What we know now is that our question should be more focused on the human side of the interaction. Students, like many of us, have learned to interact with search engines using a kind of "one-shot" query. That skill carries over to how we use generative chatbots, but, we argue, it comes at the cost of vastly under utilizing the contributions that a generative chatbot can make towards inquiry.

The range of skill needed for effective search engine query is arguably quite narrow. The best Googlers likely have distinctive ability in interpreting the results, but the disparity between the best and the worst keyword input is minimal. A good hammerer can drive a nail reliably with few swings. By contrast, we claim that generative chatbots are more like a loom for weaving, in an iterative process of refinement and corrective feedback. The large language models (LLMs) that undergird generative chatbots are initialized to be general when the dialogue begins, and as such will not yet be ideally situated to answer queries except those that are equally generic (e.g. "define X" or "who wrote Y"). The art of prompt engineering reflects one aspect of this point. A good prompt will provide the generative chatbot with the information, context, and instructions for how to respond in an effective way. Prompting, however, is a shorthand or subset of the skills of iterative inquiry that we find lacking.

In short, what we found in our classroom study was a skills gap, one that prevented a meaningful measure to assess the generative chatbot itself. Interestingly, however, students reported a perceived benefit of interacting with the generative chatbot, which we interpret as their recognition that there is something to be gained with the skill of iterative inquiry that they were learning (and admittedly, the instructor was learning to get better too).

In this paper we describe the experimental design for the study. We then report the development of the study from the instructor's perspective, using survey results and transcripts to corroborate those findings. Finally, we identify some challenges and propose some ways forward.

1.1 AI Ecology

It can be said the the social technological landscape of the 21st century as visualized through the framework of the accessible Internet, is that of search and service. Humans evidence this paradigm with the expectation that the Internet is a hyperspace tunnel to endless pages of published information, always available, to push and pull information to accomplish capricious needs. The ability to access a given state or modify some of its content relies on the graph's fundamental nature—asynchronous and persistent. The users' portal appears quiescence as it waits for their convenience. We describe a dispassionate utility.

The recent rapid uptake of generative AI tools forecasts an upgrade for the current low-quality human-computer interaction, suggesting perhaps a new moniker, HCC, human-computer conversation. The emerging AI ecology appears poised to transform expectations.

Generative Algorithms OpenAI's generative pretrained transformer(GPT) chatbot, ChatGPT, witnessed the fastest growing consumer adoption in history[2], arguably centering it as the gate-opening consumer general-purpose AI tool. Generative algorithms distinguish from discriminative ones by being able to combine and synthesize new material from spontaneous source selection. The developers of that generative AI tool chose not just language as the primary interface to the algorithm, but in fact *dialog.*

One may wonder, however, has a generation of asynchronous one-shot human-computer interactions habituated the human participant into eschewing dialog? Indeed, a prominent part of contemporary electronically moderated "speech" is instant messages, social media posts, and other timeless distributed *messages-in-bottles* that appear to discourage better-suited dialogue modalities.

To emphasize the distinction, a characteristic inquiry into the Internet or a product document or a reference text is a singular action, each solitary, even if the action is part of a sequence. Any iterative gain—from the most fundamental serial search, manually splitting a volume to find a page number, to the more complex qualitative search of scanning from a variety of sources to compile an objective fact—must be held as an adaptable representation in the seeking human's mind. Likewise, on the surface[1], any single response from a generative AI is no different, but refining the query with repeated prompts describes a process analogous to interpersonal dialog, evolving, refining, discovering, as each successive inquiry contains the entire series in a continuous context. The former examples can be considered, at best, internal monologue, while the later is an opportunity to engage in a true dialogue. The chat robot is an effective interlocutor.

In response to this emerging ecology, the researches conceived of this study, to explore reforms to formal education with the inevitable ubiquity of this new toolset.

2 Experimental Design of Intended Study

The background for the course in which the study took place is as follows. The class meets once a week for three hours. The instructor lectures for the last hour about the material for the next week. Students are then expected to read the material and write a draft paper that they will bring the next week. Each paper requires students to engage in some kind of philosophical exercise, such as a deductively valid reconstruction of an argument, the refinement of a definition by using counterexamples, or the articulation of enthymemes from the reading. When the class meets the following week, students bring their papers and work in small groups. Each group works towards a "master response" drawing from their individual papers. They then write up this response and submit it to the LMS (Learning Management System).

[1] It is trivial to conceal any number of iterative context loops before presenting a result, q.v. generative adversarial network (GAN) and variational autoencoder (VAE)

Over the course the experiment consists of the following. Each week there were four groups, two of which serve as the control, two as the intervention.[2] For the control groups, students were allowed to use their individual papers, in-group dialogues, and the Internet, but not ChatGPT or other LLM. For the intervention groups, students would have additional access to ChatGPT through our custom web app that we called MetaGeep, which would allow us to record the interactions as transcripts. Initially the version we gave them access to was GPT 3.5, but it become readily apparent after a few weeks that version 4 was significantly more powerful (this change is one reason the planned assessment became less meaningful). Assignment to groups was done randomly each week, with the condition that each student would have roughly the same number of opportunities to be in both kinds of groups.

In addition, we had an entrance survey to gauge students' prior familiarity with ChatGPT or other LLMs. At roughly two thirds of the way through the course we administered a survey to collect qualitative data about student perceptions of their use of MetaGeep. Questions comprising these surveys are included in the attached tables. Finally, we saved the transcripts of the intervention groups. The study was approved as exempt by the Internal Review Board at our home institution (IRB 23-154).

Our initial hypothesis was that intervention groups will have better group assignments than control groups. At the conclusion of the semester, the planned assessment was to anonymized each submission to be graded by an outside faculty member. We suspected that students would learn how to better interact with MetaGeep as the semester progressed; potentially co-evident in the development of their weekly group papers. We had thought that the rate of improvement week-by-week would be outstripped by the improvement between intervention and control groups for that respective week, but for reasons explained in more detail below, too many confounders entered into the study to warrant an assessment, particularly in light of the small sample size ($N = 11$). The biggest confounder throughout was that student AI literacy was far lower than we had anticipated.

2.1 MetaGeep

To facilitate interactions with ChatGPT, the researchers considered the procedural constraints of the available commercial webapp. The free app offered by OpenAI is a powerful demonstration of the LLM capabilities, and suitable for casual interactions, but the company places load limits on the freemium version, and denies access to the most advanced model and experimental features. Furthermore, users must create a personal account, and any data exchanged falls under the auspices of the TOS which is beyond our control.

With an expected class size of around a dozen students, the first consideration of creating a shared professional group account seemed clumsy and unworkable,

[2] There was also a fifth group that had a student that wished not to participate in the study, for which no data was collected.

and doesn't solve for our interest in anonymized data analysis allowed by the IRB agreement.

To solve these issues, we committed to creating a dedicated webapp to facilitate operations. OpenAI provides a commercial API interface to numerous pretrained models. The MetaGeep webapp uses the API in a basic python flask environment to intermediate between the client (the students' browsers), our data server, and OpenAI chat completion calls.

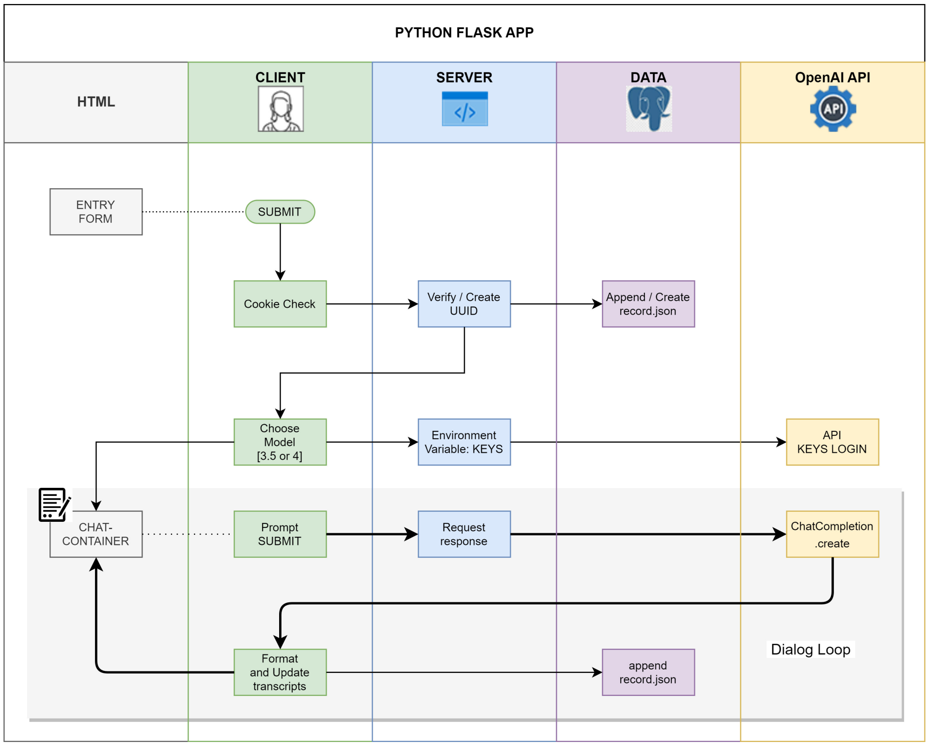

Fig. 1. MetaGeep webapp dialog loop.

Fig. 1 visualizes the webapp architecture. The client (participant student) visits the webapp via the public URL. Each access to the page requires the visitor to fill out a two question form about the group membership assignment to track minimal metadata. The client browser checks for a local browser cookie for a UUID marker, generates a new one if this is missing, and iterates a counter. A record is generated on the researchers' file server. All exchanges between the participant and the LLM is saved to a basic JSON session log file on our server.

A simple chat interface is shown to the user. At the start of the study, only models up to the GPT3.5 were available to us. After a couple months of payments, the company allows access to GPT4. This access was granted at about the mid-point of the study, after the students had taken a mid-point survey. In lieu of hard-coding a fixed connection to GPT4, we provided the users a switch in the UI that allowed them to choose the superior model. This was not necessarily out of interest of offering options (as end-users, we see no reason to spend time with 3.5 over 4), but rather as a signal to the students to increase awareness that the new model was available. Intentionally selecting GPT4 and getting suddenly better, more nuanced responses, should be done in the open.

The chat interface is a simple dialog box that expands to accommodate the transcript as it grows. It is important to note that formatting response text can take advantage of CSS styles and LaTeX symbols in the OpenAI completion payload. This feature becomes valuable when working with formal logic and mathematical symbols that are painful for humans to interpret in straight HTML text conversions. As novice webapp developers, we learned some of these minor but impactful tweaks as the study ran.

We also take this opportunity, as a topic self-referencing example, to informally share credit with ChatGPT itself, for helping write the application[3].

MetaGeep on Github https://github.com/Cognition-and-Usability-Lab/metageep

3 Results

We present the results of our observations in chronological order. By doing so we aim to emphasize the most important lesson we learned from conducting this study, which is the need to adapt to the level of AI literacy of the students. Our original study design presumed that group assignment quality could be attributed to features (or lack thereof) of the generative chatbot, even after an introduction to it through a primer exercise.

3.1 Entrance Survey

Our entrance survey provided us with information about student perceptions concerning their skills related to search and chatbots. This helped give us a baseline. The questions and responses are found in Table 1.

Our gloss of the results from the entrance survey was that students tended to self-identify as having the skills for search and for utilizing group interactions. Half the students said that ChatGPT was somewhat or more effective than search engines, and yet still lukewarm about using chatbots for learning. The inference of the two questions suggest that students find non-directed internet knowledge searches to be not very effective at all.

[3] Attributions deserved? Searching the Internet and developer forums to find samples and bug fixes is the standard for casual developers in the pre-LLM days. Today, LLMs can sketch out entire applications with functions written from scratch while we sip tea.

Table 1. Entrance Survey Responses (numbers are sums)

How familiar would you say you are with ChatGPT or other LLM?	
Not at all familiar	2
A little familiar (heard of it but haven't used it)	2
Somewhat familiar (have tried it)	2
Familiar (use it occasionally)	5
Very familiar (use it regularly)	0
How frequently do you interact with ChatGPT or other LLM?	
Never or rarely	6
At least once a month	2
At least once a week	3
At least once a day	0
Compared to search engines (e.g. Google) and information sources (e.g. Wikipedia), how would you rate ChatGPT?	
Much less effective	1
Less effective	1
Somewhat less effective	0
About the same	3
Somewhat more effective	4
More effective	2
Much more effective	0
In general, how effective do you find ChatGPT or other LLMs to be as an assistant to learning?	
Not at all effective	3
Somewhat effective	5
Effective	1
Very effective	2
How comfortable would you say you are with Internet tools like search and online information sources (Google and Wikipedia)?	
Not at all comfortable	0
Somewhat comfortable	0
Comfortable	4
Very comfortable	7
How comfortable would you say you are working with others to solve a problem or complete some task?	
Not at all comfortable	0
Somewhat comfortable	2
Comfortable	5
Very comfortable	4

3.2 Primer: What's a Sandwich?

The primer for the course is meant to do three things: 1) introduce students to the methodology of metaphysics, 2) demonstrate by example the weekly pattern we follow going forward (group work, lecture, individual paper - see above for details), and 3) introduce them to the generative chatbot they have access to for the course, which we called MetaGeep.

Metaphysics, as studied in analytic philosophy, has a very rigorous methodology. It's main tool is classical deductive logic: any argument ought to be reconstructable (in principle) into a series of deductively valid inferences. The course does not require Symbolic Logic as a prerequisite and most students will not have taken such a course. Consequently, the course begins with a logic primer.[4] In particular, the primer emphasizes the following conceptual tools that students will be most frequently exposed to and use throughout the course:

- Modus Ponens: the deductively valid inference pattern from “If A then B" and also “A" to the conclusion “B". This is contrasted with a fallacious and invalid inference called affirming the consequent: from “If A then B" and also “B" one cannot deductively infer “A".
- Modus Tollens: the deductively valid inference pattern from “If A then B" and also “not B" to the conclusion “not A". This is contrasted with a fallacious and invalid inference called denying the antecedent: from “If A then B" and also “not A" one cannot deductively infer “not B".
- A counterexample that demonstrates a purported definition or analysis is *too strong* (or *too narrow*): a case that should count as an instance of the concept but the analysis rules it out.
- A counterexample that demonstrates a purported definition or analysis is *too weak* (or *too broad*): a case that the analysis rules in but is not an instance of the concept.

In order to learn and practice using these tools, students are asked to analyze the concept of a sandwich. They are given an initial definition, such as, "A sandwich is a type of food made of two or more slices of bread with one or more fillings between them." As one might expect, students have a lot of fun coming up with counterexamples (hotdogs, calzones, poptarts, etc), updating the definition, and finding ways to defend their positions by “monster-barring”.[5]

Their first weekly individual assignment is to construct two modus tollens arguments that can be used against the initial definition.[6] The following week students are organized into groups where they are assigned to creating a “master

[4] The textbook we use in the class comes with a logic primer, which serves as a start.[5]

[5] This is in reference to the work of Imre Lakatos on the logic of mathematical discovery.[3]

[6] Specifically they are told, “In one argument (C.1) you should use a counterexample that does not count as a sandwich (per the definition), to show that the definition is too weak (includes case that should be ruled out). In a second argument (C.2) you should use a counterexample that shows the definition is too strong (excludes cases that should be ruled in).

argument" that defends an analysis of what a sandwhich is. For this first group assignment, all groups are allowed to and encouraged to use MetaGeep. The instructor provided a demonstration of how they could interact with this generate chatbot.

Previews of the transcripts from the primer suggested that students were keen to explore the use of MetaGeep as a tool by asking it questions. It was apparent, however, that students were using it just as they would a search engine: the transcripts showed a non-sequitur sequence of one-shot queries. They did not, for example, provide MetaGeep with any sort of feedback as one might in a dialogue with a peer or collaborator. This was an interesting observation because students generally did provide that kind of feedback to each other in their group interactions. We expected that as students became familiar with MetaGeep their interactions with it would more closely resemble their interactions with their peers.

3.3 Early Weeks

Reviewing the transcripts in the early weeks of the course revealed two related issues. One was that MetaGeep was using version ChatGPT 3.5 and giving "canned" style responses that did not solicit additional interactions. For example, a common way that ChatGPT 3.5 would hedge is to say things like, "different cultures and societies give different answers." This was particularly unhelpful because students are not allowed to hedge the same way in their papers and their assignments. Students must be able to explain which answers would be generated by different perspectives and why - they are not allowed to end their analyses by simply saying "it depends". We addressed this issue around week 5 by upgrading to ChatGPT 4. This was then demonstrated to all students and we illustrated how much more nuanced MetaGeep became as a result.

A second issue was that the nature of some of the assignments were too advanced or nuanced that failed to incentivize the use of MetaGeep. Specifically, some of the group assignments required students to reconstruct an argument from that week's reading into a deductively valid argument. Both the students, and the instructor, quickly faced two challenges. First, many of the topics were highly specialized and MetaGeep was not being given enough context - we were facing a prompt engineering problem. Second, the task of reconstructing an argument is not the same as providing summaries, but MetaGeep would frequently conflate these (and mind you, students often too).

In order to make some progress towards this second issue, the instructor adapted some weekly group assignments to better align with the sorts of tasks that generative chatbots seem to be helpful with. Here is one such example:

> You are going to write a short dialogue between the Presentist and the Eternalist. You can assume that both are aware of the Truthmaker Objection. Your dialogue should begin with the Presentist rejecting one of the premises. You should aim for at least two iterations (example below).

> You should also aim to have each contribution to be a concise argument. Here's an example template:
>
> Presentist: I deny premise (II) because X.
> Eternalist: Appealing to X commits you to Z. You should reject Z because Y.
> Presentist: X does not commit me to Z because W.
> Eternalist: Appealing to X commits you to U. You should reject U because V.
> Presentist: I am happy to accept U because V is not compelling.

Reviewing transcripts from this kind of assignment showed more engagement with MetaGeep relative to the "reconstruct an argument into deductive form" assignments.[7] But the kind of interaction was still largely a sequence of non-sequitur one-shot queries. By this point we were in the advanced stages of the course (week 8) and each student had been in a group with access to MetaGeep at least three times. In our assessment, the presumptions that undergirded the very ability to even test our hypothesis had failed to hold as expected in our original study design. Users have to be sufficiently proficient in using a technology if a study aims to draw conclusions about the contributions made by the technology. Consequently, we opted to use the planned exit survey to help us gauge student perspectives at this advanced stage.

3.4 Advanced Stage Survey

Our advanced stage survey aimed to gauge how useful the generative chatbot was for this course relative to other resources students had access to. Two salient questions from the survey with mean scores in Table 2.

With all the standard caveats concerning small sample sizes ($N = 11$), two points stood out to us. First, MetaGeep was, from the perspectives of students, better than other materials they could access. The exception was the instructor, who has a strong track record in teaching and has refined complementary course materials (e.g. slides, lecture notes, etc.). In jest: LLMs are not (yet) coming after professor jobs.

Second, and more pedagogically informative, is that students regard each other as having been more helpful than any other tools they had access to. While our observations do not allow us to say precisely why they have this perspective, it is consistent with the speculations we discuss in the next section. In brief, we surmise that students' ability to do iterative inquiry with their non-expert peers in a shared context is more productive than their attempt to do so with an AI that has broad expertise. But, as we will argue below, the Technological Attribution Error makes any inference particularly difficult to assess given nascent AI literacy, especially in the context of inquiry.

[7] By "more engagement" we mean a larger number of queries to MetaGeep.

Table 2. Advanced Survey Responses, Mean Scores

Rate the usefulness of these tools on a scale of 1-5, with 5 being the most and 1 the least useful.	
Internet	1.78
ChatGPT	2.22
Books, articles	1.33
Lecture materials, notes	3.33
How much do you agree with the following claims, where 1 means "Strongly disagree" and 5 means "Strongly agree"?	
Discussion with the people in my group helped shape my conclusions	3.78
The tools we used were helpful in shaping the group's conclusions	3.00
I learned or relearned somethings from the tools we used.	3.67

4 Discussion

As a gloss, LLM AI was effective for helping students summarize positions, provide a rough overview of debates, and to a limited extent helped them think through counterexamples. It seemed not to be effective for assignments that required careful argument reconstruction based on more nuanced points discussed in the text and in class. Early work elsewhere is providing preliminary evidence that that the capabilities of AI are creating a "jagged technological frontier" where some tasks can easily be done by AI, but other similarly difficult tasks are outside its capabilities.[1]

We believe that our observations over the course of this study corroborate what many early adopters are experiencing. In what follows, we present an argument that AI literacy is too nascent to draw any substantial conclusions about either this emerging technology or the users for which they are intended.

We then conclude with a suggestion for how to adapt iterative methods of inquiry, which are exemplified by, though not limited to, humanities scholars. In brief, we suggest a fruitful path forward is to guide the development and use of generative chatbots in a way that aims to elevate existing forms of iterative inquiry.

4.1 The Technological Attribution Dilemma

There are two well-known cognitive biases worth highlighting to set the stage. When someone else makes a mistake, individuals are likely to attribute that mistake to the person's character or personality ("they're careless"). However, when they make a mistake themselves, they're more likely to blame the circumstances or external factors ("the instructions were unclear"). This is the Fundamental Attribution Error, which highlights a bias in social perception, where there is an imbalance in how we interpret our own behavior versus that of others.

Relatedly, when things go well, people tend to credit their own abilities, efforts, or characteristics ("I got a good grade because I'm smart and I studied hard"), but when things go poorly, they are more likely to blame the situation, luck, or the actions of others ("I got a bad grade because the test was unfair"). This is known as the Self-Serving Bias.

There is a third related error, we suggest, that is made particularly salient with the recent advent of generative AI. We call it the Technological Attribution Error. We'll illustrate it with a mechanical example. Suppose there is a single car crash in the Town of Competence. If the baseline is that every driver in the Town of Competence is an expert driver, it would be reasonable to attribute the car crash to some property of the car - perhaps the brakes failed, for example. If the same car crash were to happen in the Town of Maladroit, where the baseline is that drivers are, at best, amateurs that couldn't even drive a go-cart, it would *not* be reasonable to attribute the car crash to a property of the car. Rather, it's far more likely that a driver was at fault.

The Technological Attribution Error occurs when someone faults a poor outcome to the technology without giving proper consideration to the skills of the operator. If we don't know whether we happen to be in the Town of Competence or the Town of Maladroit, it would be fallacious for us to infer that a car crash is the result of something about the car (i.e. the technology). Likewise, it would also be fallacious to simply attribute features to the driver (i.e. the user).

Relatedly, an Attribution Dilemma occurs when there is insufficient information about the capacities of a technology that prevents us from assessing user skills. If we don't know the handling capacities of cars (braking, turning, impact absorbing, etc) we don't have the means to conduct driver tests, let alone driver education. When we have a fuller understanding of what a vehicle can and can't do we can then develop a more systematic program for training and testing.

We unexpectedly found ourselves in an Attribution Dilemma in our first experience with using generative AI in the classroom. More sophisticated studies outside of academia suggest that we are not alone.[1] We see this as both a cautionary tale and as an opportunity. As a cautionary tale, the very concept of AI literacy is still too nascent for us to be confident in attributing general performance (both successful and unsuccessful) to just one of the user or the generative AI. As an opportunity, we are just beginning to explore the purposes for which LLMs and other generative AI can and should be put to use. In the last section we describe a possible avenue forward in the context of education.

4.2 Iterative Inquiry

We argue that the general population's accustomization to lexicographical search, at least insofar as represented by undergraduate students, makes them particularly prone to the technological attribution error when it comes to generative chatbots. That is, students are likely to engage with ChatGPT with one-shot inquiries, such that when they then receive results contrary to their hopeful expectations, they are likely to infer that the generative AI tool is of little help. What students don't seem to recognize is that they are residents in the Town of Maladroit. Consequently, students are too quick to shy away from the technology because of a fallacious inference they make about its quality. In many cases faculty are also subject to making this error.[8]

We suggest that one possible way forward is to give students assignments that better mirror the iterative nature of generative chatbots. Our suggestion is not meant to replace certain pedagogical techniques, nor even to augment them, per se. Our goal is to help think through ways in which generative AI can elevate humanistic inquiry.

To that end, we briefly describe methods of iterative inquiry. Humanities scholars in particular are experts in using iterative methods of inquiry that are directly applicable to AI literacy. Iterative methods in inquiry include the Socratic method and the method of reflective equilibrium.

The Socratic method is a form of cooperative argumentative dialogue that stimulates critical thinking and illuminates ideas through asking and answering questions. It involves a process of questioning to expose contradictions in one's thoughts and ideas, leading to clarification or reevaluation of beliefs. It is illustrated by the classic and famous dialogues of Plato.[9] It is also illustrated in more contemporary form in mathematical debates, as in the development of proofs of Euler's formula.[3]

The other related example is the method of reflective equilibrium, or more aptly called the reflective equilibrium process.[7][10] In a simple characterization, one begins with a first draft of a principle that accounts for two clear cases; one set includes instances of the relevant concept, the other not. Above we discussed a cartoonish example of this, where we have a draft definition of what a sandwich is, supplemented with some clear examples of sandwiches (a BLT, a grilled cheese sandwich) and clear non-examples (a bowl of cereal, a salad). One then proceeds through a process of mutual adjustments to our judgements about relevant cases (e.g. a hotdog) and the systematic principles meant to account for our judgments. By modifying either when they conflict, one increasingly balances judgments with general principles, ideally obtaining the goal of reaching a state of equilibrium (consistency and harmony)

4.3 Implementation

It may appear that we set the blame of underwhelming results exclusively on the shoulders of the users—for being unimaginative, uninsightful, perhaps dispassionate. As we observed the first interactions between the students with our chatty general purpose knowledge machine, we were admittedly perplexed how

[8] There is also the counterpart to this error, which is that students simply copy and paste the first result they get. This too reflects their unawareness that they are a citizen of the Town of Maladroit.

[9] See any of *Apology, Crito, Euthyphro, Phaedo, Protagoras, Meno, Symposium, or Gorgias.*

[10] It was made famous by John Rawls in his theorizing of the concept of justice.[6] The end goal of the process is typically to provide a means of justifying a theory of a concept, like justice or the precautionary principle. But that aim is not of interest here. Rather, it is the process itself that we think is relevant.

lackluster the interactions read. Keeping in mind that these are upper-division philosophy students, we expected some fireworks, as we ourselves experienced right away in each our own preliminary exposures.

We propose it is the general purpose nature of the un-tuned LLM constructs which belies the apparent disparity. Internet users of the contemporary age take their interactions for granted. The internet is on the whole, unimpressive and predictable. The advances of the the 30 or so years of the internet-connected age has focused on distribution and quick pay-offs. A young adult in 2024 developed her *theory of mind* in a world with globally-connected pocket computers, responding to her touch, video-calls with her grandmother, texting her like-wise enabled friends. To her, the inquiry mode of the Internet, is keyword search, the services are commercial and transactional, the entertainment options are broad and served without delay. Yet the intellect is dim; peak 2020 internet, to the common user, is serviced by pseudo-aware algorithmic social media trackers. Advertising.

As OpenAI proudly rolled out its shiny new language robot, they were necessarily constrained to set it into the firmament *without constrains*[11]. This was a tool for all users, for all purposes. It is an Oracle. The inference potential of any simple question, when posed to the Oracle, can map out an entirely original response as thorough as a pages-long encyclopedia entry. This is more-or-less the problem, or a problem, with LLMs as tutors. The opportunity for engagement is forestalled when the genius know-it-all gives a perfectly composed essay as reply, all in platonic Oxford English diction.

If venture capital investments are any indication of the direction that AI will be headed[4], then it appears to us that the interests in AI is corporate-profit centered. This is not on its face a bad thing, but we feel it does herald more of the same immediate pay-off driven attention, which rarely prioritizes the education sector. This is to say, the inclination for AI implementation is productivity, not well-being. While the later may follow the former, those of us heeding this attention gap may take this as a call to action.

To spell out the critic, simply attaching a flawless natural language engine onto a neural network weighted with 2 trillion parameters isn't adequate interface. The network has ingested everything it can get its hands on, pattern matching into elaborate multi-dimensional webs, and it rests waiting for your inquiry, tireless, without desire. Yet it knows comparatively nothing about the interaction it is having with its user right now. In our opinion, it needs a little more sensitivity, and restraint, and to be more inquisitive. Human-computer conversation, HCC.

[11] To be literally accurate, specific topic and behavioral constrains were always in place. Lessons learned from Tay[8].

References

1. Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R.: Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper (24-013) (2023)
2. Hu, K.: ChatGPT sets record for fastest-growing user base - analyst note | Reuters, https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/
3. Lakatos, I.: Proofs and refutations: The logic of mathematical discovery. Cambridge university press (2015)
4. Metinko, C.: The Biggest Of The Big: AI Startups Raised Huge — These Were The Largest Deals Of 2023 (Dec 2023), https://news.crunchbase.com/ai/biggest-ai-startups-openai-msft-eoy-2023/
5. Ney, A.: Metaphysics: an introduction. Routledge (2014)
6. Rawls, J.: A theory of justice. Cambridge, Massachusetts : The Belknap Press of Harvard University Press (1971), https://search.library.wisc.edu/catalog/999472448502121
7. Rechnitzer, T.: Applying Reflective Equilibrium: Towards the Justification of a Precautionary Principle. Springer Nature (2022)
8. Schwartz, O.: In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation - IEEE Spectrum (Jan 2024), https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation