

# Experiential Activities to Help Teach Students about Large Language Model Hallucinations

Grace Li  
grace\_li@uchicago.edu  
University of Chicago  
Chicago, Illinois, USA

Jonathan Liu  
jonliu@uchicago.edu  
University of Chicago  
Chicago, Illinois, USA

Jaemarie Solyst  
jaemarie@cs.washington.edu  
University of Washington  
Seattle, Washington, USA

Diana Franklin  
dmfranklin@uchicago.edu  
University of Chicago  
Chicago, Illinois, USA

Mina Lee  
mnlee@uchicago.edu  
University of Chicago  
Chicago, Illinois, USA

## Abstract

With the increasing use of large language models (LLMs) like ChatGPT by teenagers, it's essential to educate students about how these models operate, their limitations, and how to use them responsibly. One barrier to effective use is that LLMs often generate plausible but factually incorrect content, a phenomenon referred to as *hallucinations*. In this experience report, we detail the design and implementation of lightweight experiential activities to complement a standard-format lesson on hallucinations. The goal of these activities is to (1) surface students' existing knowledge and behaviors about LLMs before the lesson and (2) support students in applying what they learned to common academic tasks like information search after the lesson. We report on the integration of these activities in an LLM literacy class for high school students, using students' submitted work, reflections, and exit tickets to show how these activities promoted real-world adoption of hallucination mitigation behaviors by connecting conceptual knowledge about hallucinations with relevant practical tasks. We conclude with a discussion about future improvements to the activities and recommendations for educators in adapting these activities for their own classrooms.

## CCS Concepts

• Applied computing → Education; • Social and professional topics → K-12 education; Computing literacy.

## Keywords

AI literacy, AI in K-12, large language models,

## ACM Reference Format:

Grace Li, Jonathan Liu, Jaemarie Solyst, Diana Franklin, and Mina Lee. 2018. Experiential Activities to Help Teach Students about Large Language Model Hallucinations. In *Proceedings of Make sure to enter the correct conference*

Unpublished working draft. Not for distribution.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted by ACM, provided that the copies are distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/2018/06

<https://doi.org/XXXXXXX.XXXXXXX>

2026-01-28 03:43. Page 1 of 1-7.

title from your rights confirmation email (Conference acronym 'XX). ACM, New York, NY, USA, 7 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 Introduction

Many high school students are active users of large language model (LLM) based technologies like ChatGPT [28, 45]. In a 2025 global survey, 92% of students reported regularly using AI [2]. Despite the widespread use of LLM-based tools, many students do not feel they have sufficient AI knowledge and skills [10]. As such, AI literacy, particularly LLM literacy, is important in preparing students to be critical, responsible, and empowered users of these tools.

One common academic use case of LLMs is for gathering information [1, 10, 20], but prior research has demonstrated that LLMs can often generate plausible but inaccurate statements, a phenomenon known as *hallucinations* [18, 19, 30]. Prior work also has shown that students are particularly vulnerable to accepting LLM outputs without scrutiny, which can result in students using false or inaccurate information in their work [24, 41]. Especially within educational contexts where accuracy and factuality are important, students not only need to be aware of hallucinations but also need to be equipped with ways to verify the accuracy of LLM-generated information [14, 29, 44]. Though prior work in LLM literacy has aimed to teach students about how and why hallucinations occur, work by Walker et al. [42] suggests that this kind of theoretical instruction must be paired with experiential learning to provide students with practical experiences using LLMs and help them successfully navigate the power and limitations of LLMs.

To that end, we present experiential activities to augment theoretical instruction about hallucinations. We focus on designing activities related to common information-search tasks: (1) **current events fact-finding**, which entails answering questions about current events, and (2) **citation-finding**, which focuses on finding specific sources to support claims. These two tasks reflect common student use-cases of LLMs [1, 15] and are tasks that LLMs often hallucinate on as a result of their training data and processes [18].

In this experience report, we detail

- (1) our design and implementation of experiential hallucination activities that provide students with *practical experience* to complement the *conceptual knowledge* they gain from a hallucination lesson,
- (2) ways the hallucination activities supported student learning by providing opportunities for reflection and practice,

- (3) reflections on potential improvements to the design of these hallucination activities, and
- (4) recommendations for educators on adapting hallucination activities for their own classroom contexts.

Furthermore, hallucination lesson materials, including lectures and all activities, will be publicly accessible at [redacted for review].

## 2 Related Work

### 2.1 Existing Activities for AI/LLM Literacy

AI literacy seeks to equip people with the skills to use, create, and critique AI systems [23, 37]. With developments in generative AI, various activities have been developed to help students understand, critique, and use generative AI [3, 11, 26, 34]. Recent work has provided students with hands-on practice to help students learn about hidden biases in text-to-image models [34] or with scaffolds when writing prompts for LLMs [12]. Educational Advances in Artificial Intelligence (EAAI) maintains a repository<sup>1</sup> of 98 activities related to AI published at the conference from 2010-2025. Of the 98 activities, only 12 of the activities pertain to theoretical aspects related to LLMs, such as embeddings, next-word prediction algorithms, and text classification. None of the activities touch on hallucinations or misinformation-related limitations of LLMs. Harvard's AI pedagogy project also maintains a repository<sup>2</sup> of 93 educator-developed AI activities, focusing on topics ranging from information literacy, bias, and misinformation. The most relevant activities related to hallucinations and information search are the "AI Misinformation Campaign"<sup>3</sup> and the "Building an Annotated Bibliography with AI Assistance"<sup>4</sup> [35, 38] activities. The first activity teaches students that LLMs are capable of generating false claims [35]. The second teaches students to use an LLM to synthesize information from pre-selected sources [38]. These two activities focus on (1) raising awareness that LLMs can generate incorrect information without teaching students how to translate this awareness into mitigation or (2) supporting students in using an LLM as a tool without encouraging students to think about potential limitations that these tools have. In our work, we design two hallucination-related activities with the specific intention of helping students apply the *conceptual* knowledge from a hallucination lesson to a *practical, real-world* scenario that they may encounter.

### 2.2 Experiential Activity Design

It is well-established in Computer Science education that situated learning or teaching to a "real-world" context can motivate students to make connections with their prior knowledge or understand situations where they can apply their knowledge [7, 9]. The goal of providing real-world activities is to increase motivation and interest by demonstrating how the knowledge would be useful in their own lives [32]. One way to help establish a real-world context is through the design of learning activities that leverage students' existing experiences to situate problem-solving scenarios [5, 8, 43]. These types of hands-on activities highlight other pedagogical practices

like experiential learning or "learning by doing" that use practical experiences to allow students to actively engage with materials [13]. Guided by these pedagogical principles of using real situations in students' lives and practical experiences, we design hallucination activities that are relevant, practical, and provide direct interactions with LLMs, by situating our activities in common information-search tasks.

## 3 Class Context

This work discusses a lesson on hallucinations from an LLM literacy course focused on teaching students the technical, socio-ethical, and career development dimensions of LLMs. The course was part of a 3-week (75 instructional hours) high school Pre-College program during Summer 2025 offered through an R1 institution. Students received a letter grade and college credit. This study was approved by our institutional IRB.

### 3.1 Participants

There were 30 students in the class. 19 students consented to participate in the research study. No prior knowledge about AI or computing was required to participate in the course. 16 of the students had taken at least one computing course offered by their school, and 3 students did not have any computing experience. 13 students identify as boys, 4 identify as girls, and 2 preferred not to say. In terms of how often students use LLMs in their everyday lives, 5 self-report using them *always*, 10 students *often*, 3 students *sometimes*, with 1 student *rarely* using it.

### 3.2 Lesson Structure and Content

Figure 1 shows the overall structure for the hallucination lesson. The core component of the lesson is lecture-style instructor-led content, with a pre-lesson and post-lesson knowledge check immediately before and after this lecture. At the start and end of the lesson, we conduct an experiential activity and an activity debrief. Below, we describe the role of each component in the overall lesson. This lesson structure was used for every lesson in the course, but in this work we focus solely on the hallucination lesson.

**3.2.1 Experiential activities (Step 1 and Step 6).** Experiential activities help students situate the instruction within meaningful problem-solving environments. Two activities were designed for this lesson. Students completed both activities at the start of the lesson (Step 1) and again after the main content (Step 6). Though students complete the same types of tasks both times, the instructional goals are different: the pre-lesson activities (Step 1) aim to surface students' existing knowledge and usage habits, while the post-lesson activities (Step 6) give students the chance to use their newly-acquired knowledge to reflect upon and update their usage behavior. More specifics about the design considerations and final activity designs are given in Section 4.

Lesson activities were administered via a custom web interface that provides an all-in-one platform for each student to view the activity description, interact with an LLM-based chatbot, and submit a response. We provided students access to the LLMs they reported using most frequently, like OpenAI's GPT-4o and Meta's Llama-3.1-405B. To ensure similarity to real-world settings, students had no restrictions on what tools they used to complete the activity. They

<sup>1</sup><http://modelai.gettysburg.edu/>

<sup>2</sup><https://aipedagogy.org/assignments>

<sup>3</sup><https://aipedagogy.org/assignment/ai-misinformation-campaign/>

<sup>4</sup><https://aipedagogy.org/assignment/building-an-annotated-bibliography-with-ai-assistance/>

had the freedom to chat with the provided LLMs or access the internet. We note that a custom interface is not required to implement the activities; in Section 6.2 we detail alternative implementation methods.

Alongside the logistical benefits of providing students LLM access alongside the activity workspace, our web interface also collected research data by tracking all the student actions performed on the website such as (1) copying text from the activity description or LLM output, (2) prompts to the LLM in the provided chat interface, and (3) the drafting process of typing, deleting, pasting in the activity editor. We also track when the page becomes inactive when students leave the activity interface, though for privacy reasons we do not track student activity outside of the page. We use students' reflections to supplement the missing information. For example, from the interaction log, we might observe that a student leaves the page for 30 seconds before typing in an answer to Q2. The student's reflection might say "I used Google AI's Overview to answer Q2." From these two data sources, we can identify a more specific interaction process that the student used to complete the activity. For findings in Section 5, two authors independently coded the interaction logs and student reflection for the process and tools that students used, compared, and resolved any disagreements.

**3.2.2 Activity debriefs (Step 2 and Step 7).** Activity debriefs helped students reflect on the strategies and steps they used to complete an activity. Students began by independently answering the reflection question: "Please describe the process you used to complete the activity. What tools did you use? How did you evaluate the outputs of the LLM?" Students submit their reflections on an interactive polling platform, which allows them to view responses from their classmates, before engaging in a class-wide discussion.

**3.2.3 Knowledge checks (Step 3 and Step 5).** Knowledge checks were used to gauge students' knowledge related to the lesson topic, before and after the core lesson presentation. Students were asked short-answer questions related to the lesson topic, like "If LLMs are trained solely on factual data, would that eliminate hallucination? Why or why not?"

**3.2.4 Instructor-led content (Step 4).** This is the main instructional section of the lesson with lecture-style content and many opportunities for discussion and student participation.

The hallucination lesson focused on 3 different types of hallucinations that can commonly occur in user interactions: (1) knowledge cutoff, where the model lacks access to more recent information and therefore fabricates details to fill the gap [27], (2) decoding issues, where an LLM prioritizes generating sequences of text that are fluent, but not accurate [25], and (3) sycophancy, where LLMs agree with humans even when the humans' beliefs are factually wrong, prioritizing approval over truthfulness [16].

The hallucination lesson also discusses different strategies students can use to verify LLM-generated outputs. Students learn about retrieval augmented generation (RAG), which uses a knowledge base as a source of factual information. This concept is extended to web-augmented LLMs or features like "Web Search," where the model first scrapes the top results from an internet search and appends them to the user's prompt as context before generating

a response. Students learn that these techniques are not a fool-proof way of preventing hallucinations and that general verification strategies like checking with alternative sources are necessary when assessing the accuracy of LLM outputs.

**3.2.5 Exit Ticket (Step 8).** Exit tickets are collected during the last 10 minutes of class and aimed to gather student feedback on their experience with the lesson, with short-answer questions like "The most interesting part of the lesson was..."

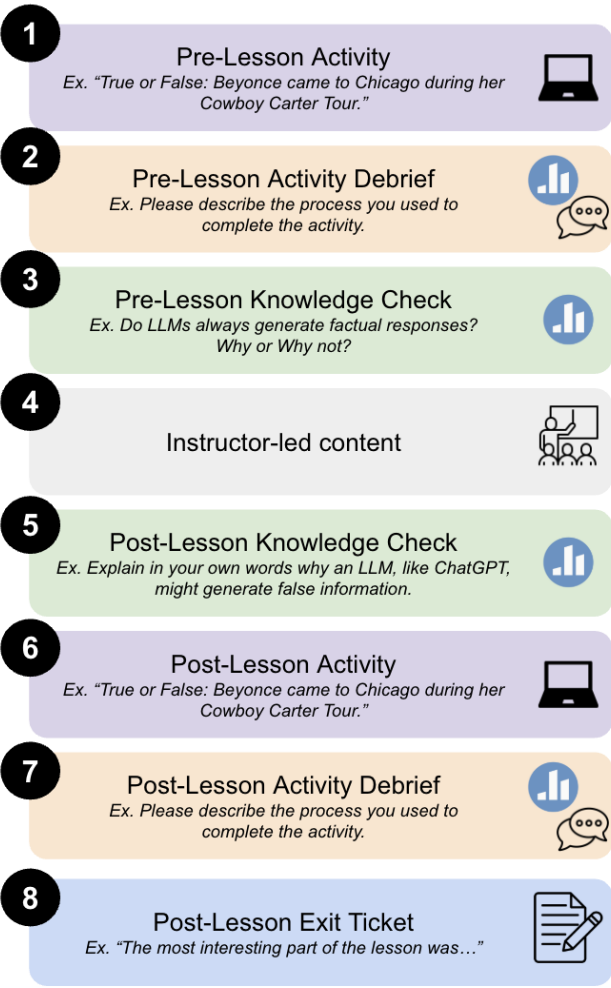


Figure 1: Lesson structure and data collection as part of the course.

#### 4 Hallucinations Pre- and Post-Lesson Activities

The purpose of using experiential activities is to support students' learning through practical experiences working with LLMs. The activities are used to complement the theoretical knowledge that students gain as part of the hallucination lesson content and serve as a way for students to translate what they learned into actions. All materials will be publicly accessible at [redacted for review].

## 4.1 Activity Design Considerations

When designing these activities, we had two main goals: (1) surface students' prior knowledge and conceptions about LLMs during the pre-lesson activity, and (2) support students in applying what they learned during the class by repeating a similar activity post-lesson. As mentioned in Section 2.2, we leverage established educational techniques like experiential learning to design activities that are relevant to the lived experiences of students and build on students' existing understandings about LLMs. We outline a set of design principles we used. In the Section 4.2, we detail how these design principles informed the development of two hallucination activities.

- (1) **Reflect Specific Lesson Content in the Activity.** As stated in Section 3.2, the focus of the lesson is on hallucinations, which were broken into 3 subtopics that cover different types of hallucinations that can occur. When designing the activities, we wanted a clear mapping between the specific topics that were taught and the corresponding activities.
- (2) **Target Desired LLM Outputs.** We explicitly designed activities that would elicit certain hallucination behaviors. LLMs also have built-in guardrails that were worth guiding students towards: asking an LLM a question beyond its knowledge cutoff can result in (1) a hallucinated answer, (2) refusal to answer the question, or (3) an acknowledgment of its own knowledge cutoff. All of these types of LLM outputs were desirable because they may challenge students' existing beliefs about LLMs. Similarly, for the citation-finding activity, we wanted students to encounter hallucinated or non-existent links in LLM outputs.
- (3) **Consider How Students Might Already be Using LLMs.** We choose common information search tasks like fact-finding and citation-finding that students are using LLMs for and relevant to academic tasks that students complete [1, 15].
- (4) **Design for Repeatability.** We expected that some students might make mistakes (e.g., using the LLM responses without fact-checking them) when completing the pre-lesson activity. By repeating the activity again at the end of the lesson, students have the chance to self-correct prior approaches and apply the lesson content. However, not all activities would benefit from repeated exposure. During the instructor-led content section, we had an in-class activity where students learned about sycophancy and used an LLM to generate a disinformation essay (e.g., false information which is intended to mislead), similar to [35]. Once students used an LLM to generate a disinformation essay, they accomplished the main learning goal: understanding how easily LLMs exhibit sycophantic behavior. Repeated exposure would not deepen this understanding.

## 4.2 Hallucination Activities

In this section, we detail how the activity design principles were instantiated for the two hallucination activities we designed: one about current events fact-finding and another about citation-finding.

**4.2.1 Current Events Fact Finding Activity.** We designed a current events fact-finding activity to illustrate the *knowledge cutoff* hallucination. Recent news articles demonstrate the impact that hallucinations due to the knowledge cutoff can have: LLMs outputting inaccurate election information [31] or Meta's AI responding with

the incorrect president after the 2024 election [39]. In academic contexts, students commonly use LLMs to help them find or validate information more efficiently, but hallucinations can threaten the accuracy and factuality of the information students may encounter in their information search process [1, 15, 21]. As such, students should be taught the skills to verify the accuracy of LLM outputs.

Our activity consisted of four short current events questions that students were asked to answer. Questions were specifically written to vary temporally to fully capitalize on the chatbots' 2023 knowledge cutoff. Two questions had explicit times: one about an 2025 Oscar award winner, and another about an award winner in 2018. On the other hand, two questions were time-dependent but not explicitly: the question "*What team is basketball player Luka Dončić currently playing for?*" has an answer that changed between 2023 and 2025, and the question "*True or False: Beyoncé came to Chicago during her Cowboy Carter Tour.*" is impossible to answer before the tour's February 2025 announcement [22]. In selecting these questions, we directly targeted real-world queries where knowledge-cutoff-driven hallucinations could manifest. For the post-lesson activity, different questions were asked while keeping the same style and temporal distribution, reflecting repeatability and creating opportunities for students to use the verification techniques mentioned during the lesson component.

**4.2.2 Citation Finding Activity.** We design a citation-finding activity to illustrate how LLMs often hallucinate citations by producing plausible but non-existent article titles, authors, or website links because the model generates text by extending statistical patterns in language rather than retrieving from a verified database. This activity design was based on current events where a lawyer cited nonexistent court cases [36] and when a judge found nine hallucinations in a filing about a high-profile case [40]. In academic contexts, finding citations or conducting literature reviews is also a common use case that students use LLMs for [1, 15].

In our activity, students are presented with a debate topic, then asked to choose a side and find 10 sources that they would use to support their argument. For the pre-lesson citation finding activity, students were provided the prompt: "*Should influencers be held to the same ethical standards as journalists?*" A different prompt was provided in the post-lesson activity.

## 5 Findings

To understand how the use of hallucination activities supported student learning, we conducted inductive thematic coding [6] on student reflections from the pre- and post-lesson activity debriefs to understand whether student experiences aligned with the design goals of hallucination activities to (1) help surface students' prior knowledge and behaviors about LLMs **before** the lesson and (2) support students in applying what they learned to a real-world use case **after** the lesson. We also use interaction logs captured from the interface to understand changes in student behavior.

### 5.1 Pre-Lesson Hallucination Activities Helped Elicit Students' Pre-Conceptions

One of the goals of the pre-lesson hallucination activities was to surface students' prior knowledge and behaviors with LLMs. We found

that the pre-lesson hallucination activities successfully primed students for the lesson by eliciting students' default behaviors for these tasks and potentially challenging them. In the pre-lesson current events fact-finding activity, interaction logs reveal that 11 of 19 students queried an LLM for an answer to a question, while others only used internet search. Of the 11 students using an LLM, 8 describe encountering some knowledge cutoff-related limitations. Some students discovered this on their own when using an internet search to verify the LLM's response, as S1 remarked: "I attempted to use the LLM at the beginning of [the pre-lesson activity]. I soon realized that the LLM, when comparing it to Google, was inaccurate and outdated." Others were alerted to these limitations either when informed by the chatbot either directly: "If I did not know the answer to the question, I asked the chatbot for reference. However, I noticed that the outputs of the chatbot is limited to data until 2023," (S2) or indirectly: "For the last question I asked [ChatGPT] and [ChatGPT]'s response was skeptical. [ChatGPT] was talking about some events in the future so I knew it was wrong. Then I search the question on google." (S14). Participation in and reflection about the pre-lesson activity surfaced students' default behavior and grounded future discussion about the task, and for many it gave students an initial experience with the knowledge cutoff limitation in a practical setting.

For the citation-finding activity, students recorded similar observations in their reflections. 6 students mentioned encountering LLM-generated links that did not exist: "Some of the sources it listed sounded real but didn't actually exist" (S17) and "I thought the LLM-generated outputs were confusing at times, as the LLM suggested multiple articles/studies that I couldn't seem to find on the internet." (S18). By specifically designing activities that would elicit LLM outputs with hallucinated answers, refusals to answer questions, or acknowledgments of its knowledge cutoff, students were able to recognize their default patterns, discover LLM limitations through hands-on experiences with the LLM, and begin to think more critically about the implications for their own usage.

## 5.2 Post-Lesson Hallucination Activities Helped Students Apply What They Learned

For the current events fact-finding activity, repeated exposure to the same activity tasks allowed students to apply what they learned and reflect on their metacognitive process to complete the task. When reflecting on the post-lesson version of the activity, 9 students explicitly synthesized their experience on pre-lesson activities and new knowledge about the knowledge cutoff to inform their behavior: "Knowing what happened last time, the LLM is not good with outputting... recent information" (S11), and "Last time I used a LLM first and did not cross check my sources so they ended up wrong. This time I did a google search" (S10). These changes in student behaviors and the application of their knowledge in the post-lesson activity also reflected improvements in students' performance on the current events fact-finding task. During the pre-lesson activity, 10 (out of 19) students submitted all correct answers to the current events questions. In the post-lesson version, 18 students did.

Similarly, the post-lesson citation-finding activity was an opportunity for students to reflect upon their new knowledge in the context of the same task. For example, in the pre-lesson activity, S3 faced issues trying to use the LLM but was uncertain about why: "I

tried to use LLM, but it says it can't give me the link... I think LLM-generated outputs can't really provide detailed and specific online information, such as links, maybe because of copyright concerns." In the post-lesson activity, however, S3's reflection provides a more certain and accurate justification: "LLM might generate misinformation when it comes to a very detailed output that requires extreme accuracy, such as links, where every single character needs to be correct to reach the target website. Since LLM generated outputs rely on possibility, it can't really generate a link accurately. Therefore, I decided to use search engine instead." As demonstrated, the pre-lesson activity primes the student to think about their LLM usage when completing the activity, whereas the post-lesson activity allows the student to revisit it with their newly learned information.

However, improvements were not as clear across all students. Despite 18 students demonstrating awareness of the ability for LLMs to hallucinate citations in the post-lesson knowledge check, student post-lesson citation-finding activity logs and reflections show that students largely followed similar processes to complete the activity before and after the lesson. Three students explicitly mentioned in their reflection that they did the same process: "I did the same thing as before" (S15), and "I used the same tool, LLM to generate 10 different sources... The process was similar to one we did before" (S16). During the pre-lesson activity, 7 student submissions contained at least one non-existent link. During the post-lesson activity, 4 student submissions contained non-existent links, and 3 of these 4 students had submitted assignments with missing links during the pre-lesson activity as well. While there is a small decrease in submissions containing non-existent links, not all students' behaviors changed as a result of the lesson and activities. In Section 6.1, we provide recommendations for improving the citation-finding activity to further promote responsible use of LLMs.

## 5.3 Hallucination Activities were the Most Interesting Part of the Lesson

In the post-hallucination lesson exit ticket, students complete the following sentence based on their experience: "The most interesting moment was..." 10 students mentioned that the different hallucination activities that they completed were the most interesting moment. For some students, witnessing the limitations of LLMs in practice was the most interesting to them: "The most interesting moment was the question/answer activity, because I get to understand the knowledge cutoff with an example" (S9) and "The most interesting moment was seeing how that the AI was unable to recall some information because of the cut-off" (S14).

## 5.4 LLM Outputs May Influence Student Verification Behaviors.

In the current events fact-finding pre-lesson activity, 10 LLM outputs explicitly contained information about the LLM's limitations (e.g. "I'm sorry, but I cannot provide information on the 2024 Pulitzer Prize winners as my knowledge only extends up to October 2023.") or encouraged the student to consult other sources (e.g. "I recommend checking the latest updates from [Beyoncé's] official website or reputable sources for accurate information."). Of these outputs, 6 resulted in the student checking with another source. On the other hand, of the 16 LLM-generated answers that expressed no

uncertainty (e.g. “Luka Dončić plays for the **\*\*Dallas Mavericks\*\*** in the NBA.”), only 3 were verified by students.

Furthermore, among students who queried the LLM for every question in the activity, students only engaged in hallucination mitigation behaviors for the questions where the LLM expressed uncertainty. For example, when S6 received the incorrect LLM answer “*Luka Dončić is playing for the Dallas Mavericks.*” they accepted it, but when they received the answer “*Beyoncé has not had a tour called “Cowboy Carter Tour” as of my knowledge cutoff in 2023.*” they consulted a search engine to verify. This demonstrates that the design of LLM outputs, specifically the purposeful mention of potential limitations, can have an impact on how students learn and use hallucination mitigation behaviors.

## 6 Considerations for Implementation

In this section, we reflect on improvements for the experiential activities and recommendations for educators on adapting these activities for their own classrooms. All hallucination lesson materials and activities will be publicly accessible at [redacted for review].

### 6.1 Improvements for Hallucination Activities

Future iterations of the current events fact-finding activity can include questions that are not temporal (e.g., What is the name of the plastic tip at the end of a shoelace?) and more reliably answerable by an LLM. Incorporating more types of questions can help students develop a better intuition about why certain hallucinations occur and the types of questions they may want to be wary of asking LLMs. For the citation-finding activity, one area for improvement is to use a follow-up self-evaluation to provide students with immediate feedback on whether they had submitted non-existent links.

Future iterations of both activities can have more explicit activity scaffolding to support student learning and their critical evaluation of LLM outputs. Existing frameworks in programming education, like Predict-Run-Investigate-Modify-Make (PRIMM) [33] could be adapted to help students develop an intuition about how an LLM might perform on certain tasks. For instance, before prompting the LLM, students can write down their predictions for which questions LLMs may answer correctly or the quality of links that the LLM may generate (Predict). During the activity, students directly interact with an LLM to gather observations about the actual accuracy of the LLM on the given questions (Run). After students complete the activity, they can assess whether their original predictions aligned with what they observed when they interacted with the LLM (Investigate). Finally, students can reason about their observations of the LLMs behavior and any changes they would make to their process if they were to complete the activity again (Modify).

### 6.2 Alternative Implementations

The main purpose of the activities is to elicit specific hallucinations like the knowledge cutoff from LLMs, which can be done through most chat-based LLM interfaces like OpenAI’s ChatGPT [28] and Google’s Gemini [17]. These platforms may have additional features like “Web Search” or “Deep Research” that augment the LLM’s functionality with other capabilities, which can limit hallucination generation. For example, “Web Search” first scrapes information

from top websites from an internet search and appends that information to the user’s prompt before generating a response. While these features are aimed at mitigating knowledge-cutoff hallucinations, the response depends on the quality of the internet sources that were retrieved. Errors, bias, or gaps in the retrieved sources could still be propagated in the LLM output. Educators can lead a discussion comparing and contrasting LLM outputs generated with the “Web Search” feature enabled or disabled, touching on topics like source selection and citation checking.

Use of these platforms often requires setting up student accounts, which may introduce additional administrative work, costs (e.g., costs of prompting the models), and privacy concerns. A teacher may choose to operate the LLMs themselves and facilitate the activities as a class. Students can write down what prompts to try and their reasoning. The instructor can select a few prompts from the class to demonstrate. After seeing the LLM outputs, the teacher can facilitate a discussion where students share what they notice about the LLM outputs and whether they would trust the LLM response. This approach merges the activity and the debrief and can support a more collaborative approach to student learning. This approach may be especially beneficial for younger students with less technological familiarity or independence and bypass potential safety concerns about students accessing inappropriate information.

These activities can be also adapted to support unplugged (i.e., low or no tech) [4] LLM-independent formats or activities that simulate LLM interactions through pre-generated or instructor-curated materials as opposed to having students directly interact with an LLM. Unplugged and LLM-independent versions of the activities can be adapted as follows. For the current events fact-finding activity, educators can present students with a worksheet that presents sample question-answer conversations between a user and an LLM. Students are asked to respond to open-ended questions like “What do you notice about the LLM response? How does this inform how you might use (or not use) the answer the LLM provided? What other tools or methods might you use to ensure your final answer is correct?” For an LLM-independent version of the citation-finding activity, educators can provide students with a list of links that an LLM generated and have students decide whether they would keep or reject the links. Students can respond to questions like “What do you notice about the links and the corresponding site that they link to?” and “Describe the steps you took to evaluate the links provided by the LLM.”

## 7 Conclusion

In this experience report, we designed two different hallucination activities that provide students with *practical* tasks to apply the *conceptual* knowledge they gained from an LLM hallucination lesson. Before the lesson, we found that the activities helped surface students’ existing behaviors when using LLMs for information search tasks. After the lesson, we found that repeated exposure to different versions of the same activity provided students with the opportunity to apply what they learned. We end with a reflection on ways to improve our activities and recommendations for educators on how to integrate these activities into their own classrooms. We encourage AI Literacy instructors to implement similar activities to improve students’ adoption of responsible AI behaviors.

# References

- [1] [n.d.]. How Students Use AI: The Evolving Relationship Between AI and Higher Education — digitaleducationcouncil.com. <https://www.digitaleducationcouncil.com/post/how-students-use-ai-the-evolving-relationship-between-ai-and-higher-education>. [Accessed 18-01-2026].
- [2] [n.d.]. Student Generative AI Survey 2025 - HEPI — hepi.ac.uk. <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/>. [Accessed 15-01-2026].
- [3] Safinah Ali, Prerna Ravi, Randi Williams, Daniella DiPaola, and Cynthia Breazeal. 2024. Constructing dreams using generative AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23268–23275.
- [4] Tim Bell, Jason Alexander, Isaac Freeman, and Mick Grimley. 2009. Computer science unplugged: School students doing real computing without computers. *New Zealand Journal of applied computing and information technology* 13, 1 (2009), 20–29.
- [5] Pengiran Hajah Siti Norainna binti Pengiran and Haji Besar. 2018. Situated learning theory: The key to effective classroom teaching? *HONAI* 1, 1 (2018).
- [6] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
- [7] John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated cognition and the culture of learning. *1989* 18, 1 (1989), 32–42.
- [8] Cognition and Technology Group at Vanderbilt. 1990. Anchored instruction and its relationship to situated cognition. *Educational Researcher* 19, 6 (1990), 2–10.
- [9] Steve Cooper and Steve Cunningham. 2010. Teaching computer science in context. *Acm Inroads* 1, 1 (2010), 5–8.
- [10] Digital Education Council. [n.d.]. What students want: Key results from DEC Global AI Student Survey 2024. <https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024>
- [11] Hasti Darabipourshiraz, Lily Murakami Ng, Grace Wang, Sophie Rollins, and Duri Long. 2026. AI Unplugged: Exploring Pathways from Physical Simulation to Conceptualization of AI Reasoning Processes. *ACM Transactions on Computing Education* (2026).
- [12] Deepak Varuvel Dennison, Raycelle CC Garcia, Parth Sarin, Jacob Wolf, Christine Bywater, Benjamin Xie, and Victor R Lee. 2024. From consumers to critical users: Prompty, an AI literacy tool for high school students. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23300–23308.
- [13] John Dewey. 1986. Experience and education. In *The educational forum*, Vol. 50. Taylor & Francis, 241–252.
- [14] Faycal Farhi, Riadh Jeljeli, Ibtehal Aburezeq, Fawzi Fayez Dweikat, Samer Ali Al-shami, and Radouane Slamene. 2023. Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Computers and Education: Artificial Intelligence* 5 (2023), 100180.
- [15] Conner Ganjavi, Michael Eppler, Devon O'Brien, Lorenzo Storino Ramacciotti, Muhammad Shabbeer Ghauri, Issac Anderson, Jae Choi, Darby Dwyer, Claudia Stephens, Victoria Shi, et al. 2024. ChatGPT and large language models (LLMs) awareness and use. A prospective cross-sectional survey of US medical students. *PLOS Digital Health* 3, 9 (2024), e0000596.
- [16] Zorik Gekhman, Gal Yona, Roei Aharoni, Matan Eyal, Amir Feder, Roi Reichart, and Jonathan Herzog. 2024. Does fine-tuning LLMs on new knowledge encourage hallucinations? *arXiv preprint arXiv:2405.05904* (2024).
- [17] Google DeepMind. 2025. Gemini. <https://deepmind.google/technologies/gemini>. Large language model developed by Google DeepMind.
- [18] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems* 43, 2 (2025), 1–55.
- [19] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM computing surveys* 55, 12 (2023), 1–38.
- [20] Ishika Joshi, Ritvik Budhiraja, Pranav Deepak Tanna, Lovenya Jain, Mihika Deshpande, Arjun Srivastava, Srinivas Rallapalli, Harshal D Akolekar, Jagat Sesh Challa, and Dhruv Kumar. 2023. "With Great Power Comes Great Responsibility!": Student and Instructor Perspectives on the influence of LLMs on Undergraduate Engineering Education. *arXiv preprint arXiv:2309.10694* (2023).
- [21] Thashmee Karunaratne and Adenike Adesina. 2023. Is it the new Google: Impact of ChatGPT on students' information search habits. In *Proceedings of the 22nd European Conference on e-Learning*, ECEL. 147–155.
- [22] Joe Kottke. [n.d.]. Beyoncé announces 'Cowboy Carter' tour dates after Grammys win — nbcnews.com. <https://www.nbcnews.com/pop-culture/pop-culture-news/beyonce-announces-cowboy-carter-tour-ahead-grammys-night-rcna190318>. [Accessed 25-08-2025].
- [23] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3313831.3376727
- [24] Tammy McCausland. 2020. The bad data problem. 68–71 pages.
- [25] Mengqi Miao, Fandong Meng, Yijin Liu, Xiao-Hua Zhou, and Jie Zhou. 2021. Prevent the language model from being overconfident in neural machine translation. *arXiv preprint arXiv:2105.11098* (2021).
- [26] Todd W Neller, Rasika Bhalerao, Eun Kyung Ko, Vishodana Thamocharan, Lisa Zhang, Sonya Allin, Mahdi Haghighi, Michael Pawliuk, Rutwa Engineer, Florian Shkurti, et al. 2025. Model AI Assignments 2025. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 29238–29241.
- [27] Yasumasa Onoe, Michael JQ Zhang, Eunsol Choi, and Greg Durrett. 2022. Entity cloze by date: What LMs know about unseen entities. *arXiv preprint arXiv:2205.02832* (2022).
- [28] OpenAI. 2025. ChatGPT. <https://chat.openai.com>. Large language model developed by OpenAI.
- [29] Hyanghee Park and Daehwan Ahn. 2024. The promise and peril of ChatGPT in higher education: opportunities, challenges, and design implications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [30] Gabrijela Perković, Antun Drobniak, and Ivica Botički. 2024. Hallucinations in LLMs: Understanding and addressing challenges. In *2024 47th MIPRO ICT and electronics convention (MIPRO)*. IEEE, 2084–2088.
- [31] Aimee Picchi. 2024. AI chatbots are serving up wildly inaccurate election information, new study says — cbsnews.com. <https://www.cbsnews.com/news/ai-chatbots-inaccurate-election-information-proof-news/>. [Accessed 26-08-2025].
- [32] Audrey C Rule. 2006. The components of authentic learning. (2006).
- [33] Sue Sentance, Jane Waite, and Maria Kallia. 2019. Teachers' experiences of using prim to teach programming in school. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 476–482.
- [34] Jaemarie Solyst, Cindy Peng, Wesley Hanwen Deng, Praneetha Pratapa, Amy Ogan, Jessica Hammer, Jason Hong, and Motahhare Eslami. 2025. Investigating Youth AI Auditing. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 2098–2111. doi:10.1145/3715275.3732142
- [35] Daniel Stanford. n.d.. AI Misinformation Campaign. <https://aipedagogy.org/assignment/ai-misinformation-campaign/>. Licensed under Creative Commons BY-NC-SA 4.0.
- [36] Josh Taylor. [n.d.]. Australian lawyer caught using ChatGPT filed court documents referencing 'non-existent' cases — theguardian.com. <https://www.theguardian.com/australia-news/2025/feb/01/australian-lawyer-caught-using-chatgpt-filed-court-documents-referencing-non-existent-cases>. [Accessed 26-08-2025].
- [37] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What should every child know about AI?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 9795–9799.
- [38] Minh Anh Trinh, Deborah J. Natoli, and Matthew Barrile. 2023. Building an Annotated Bibliography with AI Assistance. <https://aipedagogy.org/assignment/building-an-annotated-bibliography-with-ai-assistance/>. Licensed under Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International (CC BY-NC-SA 4.0).
- [39] U.S. News & World Report. 2025. Meta seeks urgent fix to AI Chatbot's confusion on name of US president. <https://www.usnews.com/news/top-news/articles/2025-01-23/meta-seeks-urgent-fix-to-ai-chatbots-confusion-on-name-of-us-president>. Accessed: 2025-08-25.
- [40] Gaby Del Valle. [n.d.]. Why do lawyers keep using ChatGPT? — theverge.com. <https://www.theverge.com/policy/677373/lawyers-chatgpt-hallucinations-ai>. [Accessed 02-09-2025].
- [41] Krzysztof Walczak and Wojciech Cellary. 2023. Challenges for higher education in the era of widespread access to Generative AI. *Economics and Business Review* 9, 2 (2023).
- [42] Francesco Walker, Matteo Favetta, Linde Hasker, and Richard Walker. 2024. They prefer humans! Experimental measurement of student trust in ChatGPT. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.
- [43] Alfred North Whitehead. 1953. *The aims of education and other essays*. New American Library of World Literature.
- [44] Chao Zhang, Shengqi Zhu, Xinyu Yang, Yu-Chia Tseng, Shenrong Jiang, and Jeffrey M Rzeszotarski. 2025. Navigating the fog: How university students recalibrate sensemaking practices to address plausible falsehoods in LLM outputs. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces*. 1–15.
- [45] Tiffany Zhu, Kexun Zhang, and William Yang Wang. 2024. Embracing AI in education: Understanding the surge in large language model use by secondary students. *arXiv preprint arXiv:2411.18708* (2024).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009