

# 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

dfranklin@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Unpublished working draft.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 educational use is granted without fee provided that copies are not made or 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.

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/XXXXXXXXXXXXXX>

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/XXXXXXXXXXXXXX>

## 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,

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

121 Furthermore, hallucination lesson materials, including lectures  
 122 and all activities, will be publicly accessible at [redacted for review].  
 123

## 124 2 Related Work

### 125 2.1 Existing Activities for AI/LLM Literacy

126 AI literacy seeks to equip people with the skills to use, create, and  
 127 critique AI systems [23, 37]. With developments in generative AI,  
 128 various activities have been developed to help students understand,  
 129 critique, and use generative AI [3, 11, 26, 34]. Recent work has pro-  
 130 vided students with hands-on practice to help students learn about  
 131 hidden biases in text-to-image models [34] or with scaffolds when  
 132 writing prompts for LLMs [12]. Educational Advances in Artificial  
 133 Intelligence (EAAI) maintains a repository<sup>1</sup> of 98 activities related  
 134 to AI published at the conference from 2010-2025. Of the 98 activi-  
 135 ties, only 12 of the activities pertain to theoretical aspects related to  
 136 LLMs, such as embeddings, next-word prediction algorithms, and  
 137 text classification. None of the activities touch on hallucinations  
 138 or misinformation-related limitations of LLMs. Harvard’s AI peda-  
 139 gogy project also maintains a repository<sup>2</sup> of 93 educator-developed  
 140 AI activities, focusing on topics ranging from information literacy,  
 141 bias, and misinformation. The most relevant activities related to  
 142 hallucinations and information search are the “AI Misinformation  
 143 Campaign”<sup>3</sup> and the “Building an Annotated Bibliography with AI  
 144 Assistance”<sup>4</sup> [35, 38] activities. The first activity teaches students  
 145 that LLMs are capable of generating false claims [35]. The second  
 146 teaches students to use an LLM to synthesize information from  
 147 pre-selected sources [38]. These two activities focus on (1) raising  
 148 awareness that LLMs can generate incorrect information without  
 149 teaching students how to translate this awareness into mitigation  
 150 or (2) supporting students in using an LLM as a tool without encour-  
 151 aging students to think about potential limitations that these tools  
 152 have. In our work, we design two hallucination-related activities  
 153 with the specific intention of helping students apply the *conceptual*  
 154 knowledge from a hallucination lesson to a *practical, real-world*  
 155 scenario that they may encounter.

### 156 2.2 Experiential Activity Design

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

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

## 174 3 Class Context

175 This work discusses a lesson on hallucinations from an LLM literacy  
 176 course focused on teaching students the technical, socio-ethical, and  
 177 career development dimensions of LLMs. The course was part of a  
 178 3-week (75 instructional hours) high school Pre-College program  
 179 during Summer 2025 offered through an R1 institution. Students  
 180 received a letter grade and college credit. This study was approved  
 181 by our institutional IRB.

### 182 3.1 Participants

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

### 192 3.2 Lesson Structure and Content

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

201 *3.2.1 Experiential activities (Step 1 and Step 6).* Experiential ac-  
 202 tivities help students situate the instruction within meaningful  
 203 problem-solving environments. Two activities were designed for  
 204 this lesson. Students completed both activities at the start of the  
 205 lesson (Step 1) and again after the main content (Step 6). Though  
 206 students complete the same types of tasks both times, the instruc-  
 207 tional goals are different: the pre-lesson activities (Step 1) aim to  
 208 surface students’ existing knowledge and usage habits, while the  
 209 post-lesson activities (Step 6) give students the chance to use their  
 210 newly-acquired knowledge to reflect upon and update their usage  
 211 behavior. More specifics about the design considerations and final  
 212 activity designs are given in Section 4.

213 Lesson activities were administered via a custom web interface  
 214 that provides an all-in-one platform for each student to view the ac-  
 215 tivity description, interact with an LLM-based chatbot, and submit  
 216 a response. We provided students access to the LLMs they reported  
 217 using most frequently, like OpenAI’s GPT-4o and Meta’s Llama-3.1-  
 218 405B. To ensure similarity to real-world settings, students had no  
 219 restrictions on what tools they used to complete the activity. They  
 220

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

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

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

4<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?"

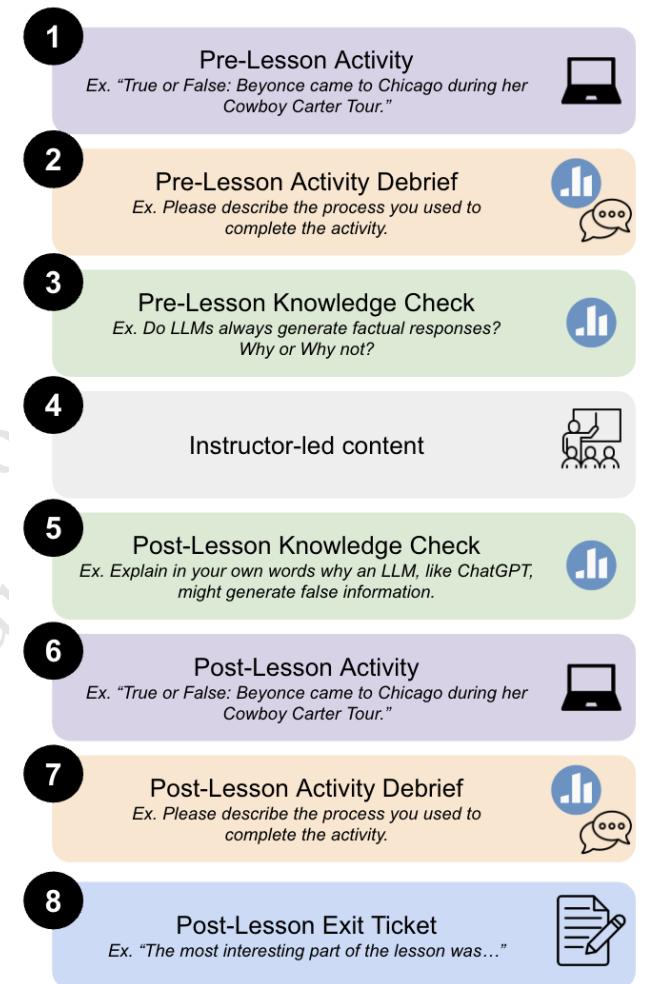
**3.2.4 Instructor-led content (Step 4).** This is the main instructional section of the lesson with lecture-style content and many opportunities for discussion and student participation.

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].

## 349 4.1 Activity Design Considerations

350 When designing these activities, we had two main goals: (1) surface  
 351 students' prior knowledge and conceptions about LLMs during the  
 352 pre-lesson activity, and (2) support students in applying what they  
 353 learned during the class by repeating a similar activity post-lesson.  
 354 As mentioned in Section 2.2, we leverage established educational  
 355 techniques like experiential learning to design activities that are  
 356 relevant to the lived experiences of students and build on students'  
 357 existing understandings about LLMs. We outline a set of design  
 358 principles we used. In the Section 4.2, we detail how these design  
 359 principles informed the development of two hallucination activities.  
 360

- (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.

## 396 4.2 Hallucination Activities

397 In this section, we detail how the activity design principles were  
 398 instantiated for the two hallucination activities we designed: one  
 399 about current events fact-finding and another about citation-finding.  
 400

401 *4.2.1 Current Events Fact Finding Activity.* We designed a current  
 402 events fact-finding activity to illustrate the *knowledge cutoff* halluci-  
 403 nation. Recent news articles demonstrate the impact that halluci-  
 404 nations due to the knowledge cutoff can have: LLMs outputting  
 405 inaccurate election information [31] or Meta's AI responding with

407 the incorrect president after the 2024 election [39]. In academic con-  
 408 texts, students commonly use LLMs to help them find or validate  
 409 information more efficiently, but hallucinations can threaten the  
 410 accuracy and factuality of the information students may encounter  
 411 in their information search process [1, 15, 21]. As such, students  
 412 should be taught the skills to verify the accuracy of LLM outputs.  
 413

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

431 *4.2.2 Citation Finding Activity.* We design a citation-finding activ-  
 432 ity to illustrate how LLMs often hallucinate citations by producing  
 433 plausible but non-existent article titles, authors, or website links  
 434 because the model generates text by extending statistical patterns  
 435 in language rather than retrieving from a verified database. This  
 436 activity design was based on current events where a lawyer cited  
 437 nonexistent court cases [36] and when a judge found nine hallucina-  
 438 tions in a filing about a high-profile case [40]. In academic contexts,  
 439 finding citations or conducting literature reviews is also a common  
 440 use case that students use LLMs for [1, 15].

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

## 447 5 Findings

449 To understand how the use of hallucination activities supported  
 450 student learning, we conducted inductive thematic coding [6] on  
 451 student reflections from the pre- and post-lesson activity debriefs  
 452 to understand whether student experiences aligned with the design  
 453 goals of hallucination activities to (1) help surface students' prior  
 454 knowledge and behaviors about LLMs **before** the lesson and (2)  
 455 support students in applying what they learned to a real-world use  
 456 case **after** the lesson. We also use interaction logs captured from  
 457 the interface to understand changes in student behavior.  
 458

### 459 5.1 Pre-Lesson Hallucination Activities Helped 460 Elicit Students' Pre-Conceptions

462 One of the goals of the pre-lesson hallucination activities was to sur-  
 463 face 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

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

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

594

## 595 6 Considerations for Implementation

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

600

### 602 6.1 Improvements for Hallucination Activities

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

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

### 629 6.2 Alternative Implementations

630 The main purpose of the activities is to elicit specific hallucinations  
 631 like the knowledge cutoff from LLMs, which can be done through  
 632 most chat-based LLM interfaces like OpenAI’s ChatGPT [28] and  
 633 Google’s Gemini [17]. These platforms may have additional fea-  
 634 tures like “Web Search” or “Deep Research” that augment the LLM’s  
 635 functionality with other capabilities, which can limit hallucination  
 636 generation. For example, “Web Search” first scrapes information  
 637

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

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

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

## 682 7 Conclusion

683 In this experience report, we designed two different hallucination  
 684 activities that provide students with *practical* tasks to apply the *con-*  
 685 *ceptual* knowledge they gained from an LLM hallucination lesson.  
 686 Before the lesson, we found that the activities helped surface stu-  
 687 dents’ existing behaviors when using LLMs for information search  
 688 tasks. After the lesson, we found that repeated exposure to different  
 689 versions of the same activity provided students with the opportu-  
 690 nity to apply what they learned. We end with a reflection on ways  
 691 to improve our activities and recommendations for educators on  
 692 how to integrate these activities into their own classrooms. We  
 693 encourage AI Literacy instructors to implement similar activities  
 694 to improve students’ adoption of responsible AI behaviors.

695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738

## 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, Prema 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 Varvelle Dennis, Ray celle CC Garcia, Parth Sarin, Jacob Wolf, Christine Bywater, Benjamin Xie, and Victor R Lee. 2024. From consumers to critical users: Promptly, 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 Shabbir 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, Roee Aharoni, Matan Eyal, Amir Feder, Roi Reichart, and Jonathan Herzig. 2024. Does fine-tuning llms on new knowledge encourage hallucinations? *arXiv preprint arXiv:2405.05904* (2024).
- [17] Google DeepMind. 2025. Gemini. <https://deephmind.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 Seth 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] Thashmeen Karunarathne 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-reca190318>. [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 Thamotharan, Lisa Zhang, Sonya Allin, Mahdi Haghigham, Michael Pawlik, 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 primus 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