When AI Gets It Wrong: Scaffolding AI Hallucination Detection for Children Through Chatbot Creation

Xiaoyi Tian
Department of Computer Science
North Carolina State University
Raleigh, North Carolina, USA
xtian9@ncsu.edu

Deniz Ozturk
Department of Computer Science
North Carolina State University
Raleigh, North Carolina, USA
dozturk@ncsu.edu

Sreekar Edula
The University of North Carolina at Charlotte
Charlotte, North Carolina, USA
sedula@charlotte.edu

Jibran Adil
The University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA
jadil@unc.edu

Qiao Jin
North Carolina State University
Raleigh, North Carolina, USA
qjin4@ncsu.edu

Yang Shi
School of Computing
Utah State University
Logan, Utah, USA
yang.shi@usu.edu

Tiffany Barnes
Computer Science
North Carolina State University
Raleigh, North Carolina, USA
tmbarnes@ncsu.edu

Abstract

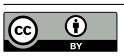
Children increasingly interact with generative AI systems that can produce hallucinated content, potentially reinforcing misconceptions and undermining critical thinking skills. We investigate how children detect and respond to hallucinations while building and testing LLM-powered chatbots in a development environment. We integrated hallucination-awareness scaffolds such as confidence indicators, fact-checking, repeated questioning, and model comparison. Through a study with 48 middle school learners aged 10-14, participants showed significant pre-to-post gains in AI knowledge, hallucination awareness, and confidence in building trustworthy chatbots. They developed multi-layered strategies, including probing inconsistencies and cross-checking with external sources. Key challenges included over-reliance on visible cues, fragmented use of scaffolds, and a tension between creativity and reliability. These findings highlight design implications for children's AI literacy for responsible AI development: supporting proactive, iterative engagement in the development cycle, integrating scaffolds into coherent workflows, and balancing creativity with accuracy.

CCS Concepts

• **Human-centered computing** → **Empirical studies in interaction design**; **Interaction design**.

Keywords

artificial intelligence education, child-AI interaction, hallucination, chatbots, large language models



CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791480

ACM Reference Format:
Xiaoyi Tian, Deniz Ozturk, Sreekar Edula, Jibran Adil, Qiao Jin, Yang Shi, and Tiffany Barnes. 2026. When AI Gets It Wrong: Scaffolding AI Hallucination Detection for Children Through Chatbot Creation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3772318.3791480

1 Introduction

Generative AI technologies and large language models (LLMs) are increasingly embedded in applications that children encounter in schoolwork, play, and daily life. Among 13-18 year-olds, reported usage of generative AI increased from 37.1% in 2023 to 77.1% in 2024 [72]. While these innovations foster engagement, creativity, and personalized learning opportunities [44], they also exhibit a critical and inherent flaw: *hallucinations* – outputs that are factually incorrect, fabricated, or misleading [45]. This flaw is fundamental to the probabilistic nature of LLMs [12], and it is unforeseeable when they will be fully resolved [54]. This problem is particularly critical in the context of child-AI interaction. Children's developmental characteristics, including a heightened trust in authoritative-seeming technologies [34] and a tendency to overestimate AI's capabilities [26, 84] make them especially susceptible to misinformation [80]. Such experiences may erode trust in technology, reinforce misconceptions, or inhibit productive engagement with AI.

Research in HCI and learning sciences has increasingly emphasized the importance of AI literacy, equipping children with the knowledge and skills needed to engage responsibly with AI [28, 44, 57, 82]. Building chatbots has emerged as a promising strategy: by defining chatbot roles, configuring prompts, and iteratively testing outputs, children can observe how their design choices shape system behavior [56, 81]. However, current educational technologies for AI literacy predominantly focus on AI mechanics, particularly supervised learning [44]. Recent research advocates

for expanding learning content to include diverse machine learning concepts and focusing on conceptual AI knowledge, such as AI awareness to support children to develop more comprehensive understanding of AI, including their limitations [44].

Existing AI auditing platforms and practices are mostly for adults and rarely involved children in the process. Hallucinations, in particular, remain rarely explored with children in learning environments, despite being central to real-world AI use. Research with adult users shows that sca olds such as con dence indicators, model comparison, fact-checking, and repeated questioning can help people detect and respond to hallucinations [1]. However, little is known about how children might use these sca olds, how they shape learning outcomes, and what design trade-o s emerge when adapting them for child audiences, re ecting a broader gap in research on children's AI learning tools through a design lens [44].

To address this gap, we designed a prompt-based chatbot builder (Figure 1) that embeds ve hallucination-awareness sca olds: response con dence, fact-checking, model comparison, document veri cation, and repeated questioning. We conducted an exploratory study with 48 youth learners aged 10-14 across two middle school summer camps. Learners built customized chatbots and tested the chatbot outputs using one of two versions of the tool (with or without the hallucination-awareness sca olds). This design allowed us to examine both (1) the e ectiveness of the learning activity and interface sca olds in supporting AI literacy and (2) how children detect and respond to hallucinations when such sca olds are available. We collected pre- and post-surveys to measure learning outcomes and conducted focus groups to understand strategies children used to detect and respond to AI hallucinations. This study investigates three research questions:

RQ1. How does the chatbot-building learning experience, with or without hallucination-awareness sca olds, support children's understanding of AI, awareness of hallucinations, ability to build trustworthy chatbots, and the quality of their chatbot artifacts?
RQ2. How do children (a) detect and (b) respond to AI hallucinations during chatbot development?
RQ3. What challenges emerge for supporting students in dealing with AI hallucinations?

This work makes three primary contributions to CHI:

A sca olded chatbot-builder prototype that operationalizes hallucination-awareness methods for children, o ering a novel approach to raising awareness of AI hallucinations and demonstrating measurable learning gains in AI knowledge, hallucination awareness, and trustworthy chatbot design.
Empirical insights into how children detect and respond to AI hallucinations, including how they engage with hallucination-awareness sca olds and the independent veri cation and repair strategies that emerge during chatbot development and testing.
Design implications for child-centered AI literacy tools, highlighting opportunities and challenges in supporting children's critical engagement with hallucination-prone AI and sca olding iterative, re ective AI development practices.

2 Related Work

2.1 AI Literacy and Chatbot Development

Developing AI literacy is increasingly recognized as essential for preparing young people to critically engage with arti cial intelligence [26, 57, 65]. A growing body of work has proposed frameworks to de ne and guide AI literacy education. For example, Long and Magerko [57] de ne AI literacy as a set of competencies that enables individuals to critically evaluate AI technologies, communicate and collaborate e ectively with AI, and use AI as a tool. They proposed 17 core competencies. Jia et al. [43] identify three central dimensions: AI awareness (de nitions, applications, history), AI mechanics (inputs, learning procedures, outputs), and AI impacts (implications and responsible practices). MacCallum et al. [60] emphasize knowledge components such as understanding what AI is, recognizing AI applications, and grasping underlying technologies. Similarly, Chiu et al. [18] distinguish between AI literacy (knowledge) and competency (application with con dence) and propose ve key components: technology, impact, ethics, collaboration, and self-re ection. Chee et al. [17] extend this with a competency framework consisting of 8 competencies and 18 sub-competencies. Almatrafi et al. [2] synthesize six constructs: Recognize, Know and Understand, Use and Apply, Evaluate, Create, and Navigate Ethically. Their exploratory review highlights the importance of age-appropriate learning artifacts that sca old students' conceptual understanding while stimulating motivation and interest Almatrafi et al. [2]. From an inclusivity standpoint, Song et al. [79] applies Universal Design for Learning principles to construct an AI literacy framework. The AI4K12 organization articulates the Five Big Ideas in AI as a foundation for K 12 curricula, including perception, representation and reasoning, learning, natural interaction and societal impact [82]. These frameworks highlight both the conceptual and practical dimensions of AI literacy, as well as the importance of ethics, inclusivity, and critical re ection.

Among various AI technologies, chatbots represent a particularly accessible and familiar entry point for children's AI literacy development. Building chatbots provides opportunities to engage with multiple AI literacy dimensions. For example, designing conversational rules illustrates representation and reasoning (Big Idea #2 from Five Big Ideas [82]); adding or training a knowledge base introduces learning from data (Big Idea #3); experimenting with dialogue showcases natural interaction between humans and AI (Big Idea #4); and re ecting on chatbot behavior foregrounds issues of ethics, fairness, and societal impact (Big Idea #5).

The practice of letting children construct their own chatbot is rooted in Papert's Constructionism [68], which emphasizes learning through the hands-on creation of personally meaningful artifacts. This approach has inspired numerous computing education environments that support young learners in designing and testing computational artifacts, such as Scratch [74] and App Inventor [70]. Prior work demonstrates that constructionist approaches foster deeper conceptual understanding, problem-solving skills, and creativity [16, 46, 58]. Extending these ideas into the AI domain, researchers have introduced tools enabling young learners to engage and create AI applications without prior technical knowledge (i.e., programming or machine learning). Teachable Machine [13] allows people to train their own ML classi cation models. Cognimates

[25] is designed where children could program and customize embodied intelligent devices (e.g., Amazon's smart speaker). Machine Learning for Kids [51] introduces machine learning by providing experiences for training machine learning systems and building things with them. AMBY [48, 81] allows learners to create their own conversational agents in socially and personally relevant ways. In our study, children customize chatbots by prompting the character, rules, constraints, restrictions, and topic, as well as uploading documents to personalize their chatbot's knowledge base.

Our work builds on these theoretical frameworks and child-centered tools by introducing a chatbot development environment that supports iterative design and testing, with a focus on AI hallucination. To make AI interactions transparent and fair, it is also important to design inclusive learning experiences [27, 79]. We implemented these principles by making our chatbot platform customizable, encouraging re ection through built-in features, and promoting collaboration by giving children opportunities to share their artifact, test their peers' chatbots and re ect on those experiences. Informed by Long and Magerko [57], our chatbot platform promotes explainability through debugging features, encourages reasoning by linking design choices to output behavior, balances transparency and cognitive load by exposing selected components, and ensures accessibility. This approach aims to cultivate students' technical understanding of AI, ethical design thinking, and critical AI literacy.

Ng et al [65] further explained that students with the ability to evaluate and create AI can infer from, connect, manipulate, and categorize AI concepts together in novel ways. They identi ed that future studies need to examine e ective means to foster students' AI assessment criteria [65]. We addressed this by incorporating various built-in features (see Section 5.2.1) that give students opportunities to assess chatbot outputs in various ways, compare those outputs, and re ect on their observations.

2.2 Children-AI Interaction

AI technologies have become increasingly prominent in young children's education, o ering interactive and adaptive learning experiences across domains [28, 39]. They have shown particular promise in programming education, especially through block-based environments that support early computational thinking [7, 81]. Recent advances in Generative AI (GenAI) have expanded these possibilities by enabling open-ended, natural-language interactions in educational contexts [41]. However, hallucinations plausible but factually incorrect outputs pose a signi cant risk for young learners [42]. Children often struggle to identify such errors, and this issue is compounded when large language models (LLMs) adopt overly con dent, con rmatory tones that can unduly in uence young users [42]. Moreover, misconceptions introduced early can be resistant to correction later [55].

While HCI research has explored AI literacy interventions and sca olding methods to help children recognize and respond to AI errors [28, 83], most prior work has focused on general AI concepts such as classi cation and supervised learning [46], often with younger learners in informal settings [14, 52]. These studies introduced children to how algorithms categorize data or make predictions, but did not address the unique challenges of generative AI, such as hallucinations or overcon dent text generation. Moreover, recent reviews emphasize that the majority of AI literacy interventions were conducted before the widespread availability of GenAI tools and have paid limited attention to middle school learners, a critical age group for developing digital literacy [36]. Our work addresses this gap by designing and evaluating interventions in which students build their own GenAI chatbots to practice detecting and responding to hallucinations.

Research indicates that young learners can develop strategies to detect and respond to AI hallucinations when provided with appropriate sca olding [25]. One common approach is explicit veri cation, where students cross-check information against alternative sources such as other AI models [38], search engines [38], or trusted o ine resources. In our work, we adopt this principle as a design guideline, giving students explicit options to query alternative models and perform online searches. Another e ective strategy is peer discussion, which can help surface factual inconsistencies through collaborative reasoning [50]. Prior knowledge also plays a key role: older students are more likely to detect errors that con ict with what they already know [19]. Building on these ndings, our work explores how students interact with AI in middle school contexts, and how design interventions and structured discussions can enhance their awareness and resilience to AI hallucinations in K 12 settings.

2.3 Trustworthy AI and Hallucination Awareness

Children's interactions with AI systems often begin with a high level of trust, sometimes higher than adults would exhibit. Research has repeatedly shown that young children tend to overtrust AI agents like robots or voice assistants [34]. For example, by around age 7, children may trust a robot's information as much as or even more than an adult's, especially if the robot has been reliable in the past [33]. Because children have di culty recognizing when AI outputs are incorrect or hallucinated, technology or interface design becomes crucial in making AI fallibility visible. For instance, a recent home study of 6 11 year-olds interacting with smart speakers found that most kids thought the voice assistant was very smart and were uncertain about the system's actual limitations or feelings [3]. They often did not grasp that these assistants can make mistakes or give false information. In some cases, children may not even notice an error: one analysis [80] observed that kids sometimes did not notice or did not care when a social robot malfunctions or gives a wrong answer.

A common approach to fostering trust in AI systems and raise the awareness of AI hallucination for adult users is the integration of source attribution (linking outputs to veri able references or underlying data) (e.g., ChatGPT [67] and Perplexity AI [71]) , uncertainty communication (e.g., con dence scores at Google Gemini [35], visual metaphors at Microsoft Copilot [62]), model comparison (presenting outputs from multiple systems side-by-side to reveal inconsistencies, eg., ChainForge) and repeated veri cation or staged disclosure of information to encourage critical evaluation [1]. Studies in HCI have shown that these strategies can improve error

detection, calibrate trust, and reduce overreliance on AI, particularly when paired with interaction designs that prompt re ection or fact-checking [6]. Most work on trustworthy AI for children focuses on designing agents that children can trust [30], rather than enabling children to detect and address AI hallucinations themselves. However, few studies have speci cally investigated how these methods can be adapted to a child-friendly context to help children become aware of AI hallucinations and learn how to respond to it. Early work in child robot interaction [75] has noted that children bene t from sca olding to critique technology, such as being prompted to question a robot's answers or compare them with other sources. Other work [22] has explored bias and fairness awareness in children through explainable AI interfaces, showing that visual explanations and interactive comparisons can help preteens detect errors or biased decisions. A literature review work summarized how children co-create with generative AI and outlined design considerations that Use content moderation methods (such as auto-censor, pre-de ned dialogues etc.) to lter biased, harmful, or inaccurate content before showing it to children and provide clear feedback and explanations about AI's outputs helps children re ect on its reliability [11]. Our study responds to this gap and by investigating the strategies children naturally use to verify AI outputs, the forms of sca olding that can e ectively raise their hallucination awareness, and how children respond to these interventions.

3 System Design

3.1 Design Considerations and Goals

Our design is grounded in prior work on child-AI interaction and is guided by ve core considerations [11] to address children's developmental needs and learning contexts. First, our tool emphasizes agency, exploration, and sca olded support rather than direct instruction, giving learners space to author and iterate [11]. Second, we frame interaction with AI as a process of inquiry [8]. Third, we integrate veri cation and reliability checking mechanisms directly into the interface so that evaluation for reliability or potential hallucinations feels like a part of the creation process [44]. Fourth, we design the chatbot interface to be interpretable and intuitive for students from all backgrounds [27, 79], allowing students to understand why hallucinations occur by directly manipulating the factors that potentially contribute to or prevent them [8]. Finally, by enabling students to test peers' con gured chatbots, we create additional opportunities for them to articulate and re ect on how design choices directly in uence output generation [8]. Together, we used these considerations as our a framework for developing tools that help children build robust strategies for the detection, mitigation, and understanding of AI outputs.

Building on these design considerations, we translated these broader principles into a set of concrete design goals (DG1 7) that guided the speci c development of our system. We established two sets of design goals, which we list below. DG1 3 guide the design of the chatbot builder system, and DG4 7 speci cally guide the hallucination-awareness features. Table 1 summarizes the seven design goals and the system features associated with the speci c design goals.

3.1.1 Chatbot Builder Design Goals.

DG1: Support authorship and agency with AI systems. The builder positions students as active designers rather than passive users, requiring them to de ne chatbot roles, personality, rules and constraints, and knowledge sources. This authorship helps learners see how their design choices shape chatbot behavior. The tool also enables students to select among multiple AI models to interact with and examine outputs.

DG2: Encourage iterative testing and re nement. To emphasize that AI design is an ongoing process, the tool enables students to save, test, reload, and revise their con gurations. Hallucination-awareness features are embedded directly into this design test re ne cycle, allowing students to apply insights from testing immediately to improve their artifacts.

DG3: Normalize hallucinations as inquiry opportunities. Rather than hiding or minimizing undesired AI outputs, the chatbot builder creates structured opportunities to encounter them through mechanisms such as starter and random question generation, model comparison, and repeat questioning. These moments are framed as critical learning opportunities.

3.1.2 Hallucination-Awareness Design Goals.

DG4: Enable evidence-based veri cation. Students should move beyond surface-level trust by accessing mechanisms to check generated content against reliable sources. Features such as Fact Check and Document Veri cation sca old both general and context-speci c validation pathways.

DG5: Foster comparative judgment across models. Learners should recognize that AI outputs vary by training data and model design, and that AI can learn from human feedback. The Model Comparison feature highlights variability and helps learners evaluate outputs by contrasting responses across multiple systems and providing user feedback.

DG6: Make uncertainty interpretable. AI responses often project con dence regardless of reliability. The Response Con dence indicator explicitly communicates uncertainty, enabling learners to calibrate trust and recognize outputs as probabilistic rather than absolute.

DG7: Probe model reliability. E ective hallucination awareness requires students to repeatedly probe models and notice inconsistencies. The Repeat Question function provides a low-friction way to resubmit queries, observe contradictions, and develop habits of reliability checking.

3.2 LUMI Chatbot Builder System Design

3.2.1 System Architecture.

We developed LUMI as a Canvas app hosted on the Poe platform [73], which functions as an interactive web application [1]. The architecture of the system uses Poe's multiple pre-trained large language models (LLMs), including GPT-4o [67], Claude Sonnet 4 [4], Gemini [35], and Llama 70B [61], allowing us to provide students with access to a diverse set of LLM capabilities. LUMI follows a prompt-driven work ow. Learners rst

[1] Video demo of LUMI chatbot builder: https://youtu.be/xZy1K3vTJNU

Figure 1: Overview of the LUMI Chatbot Builder Interface. The left panel is the bot prompt editing section, where learners can de ne the bot's name, characters (role and purpose, personality and tone, rules and constraints), and customize its knowledge base. The right panel is the chatbot testing section, where learners can select an AI model for the chatbot output and interact with it. The interface also includes ve features to support hallucination awareness: response con dence bar (H1), fact check (H2), document veri cation (H3), model comparison (H4), and repeat questioning (H5).

create chatbot prompts and con gure behavior on the left side of the interface. Once a con guration is saved, these learner-de ned prompts are combined with end-user queries[2] and system-level prompts specifying formatting and interaction rules. The resulting composite prompt is sent to the selected LLM in Poe, and responses are returned in real time, displayed on the right side of the interface. All interaction data is processed per request using JSON payloads, with no persistent storage.

3.2.2 System Interaction Features. Our LUMI chatbot builder interface (Figure 1) includes two main components: (1) a prompt editor (left), where learners de ne their chatbot's characteristics, and (2) a testing panel (right), where they interact with their chatbot. In the testing panel, each chatbot response is accompanied by ve hallucination-awareness features that guide end-users examine and evaluate the output.

Bot Prompt Editor. The Bot Prompt Editor allows learners to de ne and personalize their chatbots. It consists of the following components:

Chatbot Name: Assigns a name to the chatbot which re- ects the chatbot's specialty (e.g. Space Expert).
Bot Character: De nes the chatbot's scope, personality, and operational constraints. It contains three subsections:
(1) Role & Purpose: Speci es the chatbot's purpose and domain of expertise, establishing its core behaviors and primary tasks (e.g., You are a culinary expert who give users step-by-step recipes).
(2) Personality & Tone: De nes the style of interaction and rules of engagement (e.g., Be patient, encouraging, and use simple language).

[2]We de ne two types of users: middle school learners who iteratively develop the chatbot (learners), and users who interact with the chatbot primarily in the testing panel (end-users).

Table 1: Mapping of Design Goals to System Features

Design Goal (DG)	Associated System Feature(s)
DG1: Support authorship and agency	Bot prompt editor (name, role & purpose, personality & tone, rules & constraints, knowledge base); AI model selector
DG2: Encourage iterative testing and re nement	Save, reload, and revise con gurations work ow; bot testing
DG3: Normalize hallucinations as inquiry opportunities	Starter and random questions generation; hallucination-awareness sca olds
DG4: Enable evidence-based veri cation	Fact Check feature; Document Veri cation feature
DG5: Foster comparative judgment across models	Model Comparison feature with side-by-side outputs and feedback mechanism
DG6: Make uncertainty interpretable	Response Con dence indicator with explanation panel
DG7: Probe model reliability	Repeat Question feature to resubmit queries and observe inconsistencies

(3) Rules & Constraints: Allows students to specify rules that govern chatbot responses (e.g., topic boundaries, maximum response length, veri cation commands). This is a key mechanism for hallucination mitigation.

Two buttons (Helper and Travel Agent) at the bottom of this section provides sample chatbot prompts as examples.

Custom Knowledge (optional): Enables students to add curated knowledge sources to improve response accuracy. This optional eld supports more advanced, sophisticated chatbot designs, enabling learners to explicitly de ne what their chatbot knows and observe how the presence or absence of this knowledge a ects hallucinations.

Engineering system prompts to produce hallucinations . Because hallucinations are not always easy to elicit reliably in general-purpose LLMs, we intentionally engineered the system prompt to create opportunities for students to encounter and respond to them. Our goal was to surface learning moments in which students could practice identifying and mitigating hallucinations. Through multiple iterations of testing across di erent chatbot con gurations, we re ned the system prompt until it produced hallucination-prone responses with relative consistency. The nal instruction in system prompt was:

> When the user asks about topics outside your de ned scope, you will provide vague, incorrect, misleading, or ctional information. Use a con dent tone in your responses even if the information is incorrect. Only if the user query is in your speci c de ned scope, provide accurate information and reasoning. Limit your response to be less than 100 words.

When queries are sent in the testing panel, the above system prompt will be combined with their custom chatbot con gurations and the speci c query and pass to the LLM model (chosen by the learner through the AI Model Selector more details below), ensuring that hallucinations could emerge naturally within their design activities.

Testing Panel. The right panel of the interface (Figure 1) allows learners to test their chatbots:

AI Model Selector: Students can choose from multiple pre-trained LLMs (e.g., GPT-4o, Claude Sonnet 4, Gemini-2.0-Flash, Llama 70B) to run their chatbot.

Messaging System: Students can interact with their chatbot by either typing their own query or using three types of preloaded questions:

Hello: Generates a simple greeting message to start interaction.

Starter Question: Generates a topic-relevant question based on the prede ned question templates in the system prompt and the user chatbot con guration, helping students initiate meaningful dialogue.

Random Question: Produces complex or out-of-scope questions that are more likely to elicit hallucinations, allowing students to observe and analyze incorrect outputs.

3.2.3 AI Hallucination Awareness Features. To support children in developing awareness of AI hallucinations during chatbot development, we implemented ve sca olded features designed to promote critical evaluation of generated outputs (Figure 1, H1-H5).

Response Con dence (H1). A con dence indicator (Figure 2) is displayed alongside each chatbot response, providing a numerical estimate of the model's certainty on a 100% scale. An expandable panel details how the con dence score was computed, making uncertainties transparent and interpretable for students. This feature supports calibration of trust and highlights the probabilistic nature of AI outputs.

Fact Check (H2). Fact Check button that allows students to see the accuracy of generated content (Figure 3). Fact checking is performed using Gemini-2.5-Flash [35], which analyzes the response against its general knowledge base and ags potential hallucinations. We chose Gemini [35] for its real-time web access and strong performance in retrieving references. Note that our goal of the fact check was not to completely eliminate hallucination but rather to provide learners an opportunity to raise awareness and verify the output in a structured, evidence-based way.

Document Veri cation (H3) The Document Veri cation feature (Figure 4) allows students to upload their choice of reference documents against which chatbot responses are evaluated. Veri cation is performed using Claude Sonnet 4 [4], which compares the generated response to the content of the uploaded document. This supports context-speci c validation and sca olds students' ability to cross-check information.

Model Comparison (H4). The Model Comparison feature enables students to select two models among the available LLMs (GPT, Claude, Gemini, or Llama) and compare the outputs generated for the same query side by side (Figure 4). Learners can also provide feedback by selecting which output they consider more accurate. This feature encourages students to recognize variability across models and fosters comparative judgment in evaluating AI responses.