

## Using AI in Book Writing with Human Supervision: A Case Study in Astronomy for Kids

In recent years, the intersection of artificial intelligence (AI) and creative writing has opened new possibilities in producing educational content. This paper explores the process of co-authoring a children's book on astronomy using AI as a writing assistant, under the supervision of a subject matter expert. While AI has advanced significantly in generating coherent, relevant text, human guidance is crucial to ensure scientific accuracy, pedagogical appropriateness, and an engaging narrative tailored to young readers. In this case study, I, the principal author and co-creator, utilized AI writing tools to draft a book aimed at children aged 7 to 14, explaining key astronomical concepts in an accessible and engaging way. The book, titled "**Los Fundamentos de la Astronomía**," was framed within the context of Colombia's night sky, making it both regionally relevant and scientifically sound. The project originated from my attendance at an astronomy conference in Athens, Greece, where I met José, a Colombian astronomer who described the challenges of teaching astronomy in Colombia using textbooks written from a European perspective. The positional differences in celestial observations between Spain and Colombia created significant confusion for his students. Recognizing an opportunity to address this regional educational gap, I proposed a collaboration using AI as a tool combined with his astronomical expertise to create culturally appropriate educational materials.

### Scenario: Collaboration between AI, the Astronomer, and the Co-Author

During the writing process, the AI proposed a unique analogy to explain black holes to young readers. It suggested describing a black hole as a "cosmic whirlpool" that pulls everything, even light, into its center, much like water going down a drain. Initially, this seemed like a promising metaphor to simplify the complex concept. As the co-author, I saw its creative potential for capturing children's imagination and visualizing an abstract phenomenon.

However, when the draft was reviewed by the astronomer, concerns arose. The expert pointed out that while the analogy of a whirlpool was effective in illustrating the gravitational pull, it could mislead young readers into thinking that black holes are visible structures with a definitive center like a drain, which is not scientifically accurate. The astronomer further explained that the event horizon and singularity needed clearer, albeit simple, definitions to avoid such misconceptions.

To address these concerns, I provided the following prompt to the AI:

*"Write a simple explanation of black holes for children aged 7 to 14. Describe a black hole as something invisible that pulls everything around it, like a powerful force in space. Use the analogy of a whirlpool, but make sure to explain that we cannot see a black hole, only how it affects the space around it. Also, clarify that the black hole doesn't have a visible 'center' like a drain, but its gravity is so strong that it pulls in even light. Keep the explanation fun and engaging, without overwhelming the reader with complex scientific terms."*

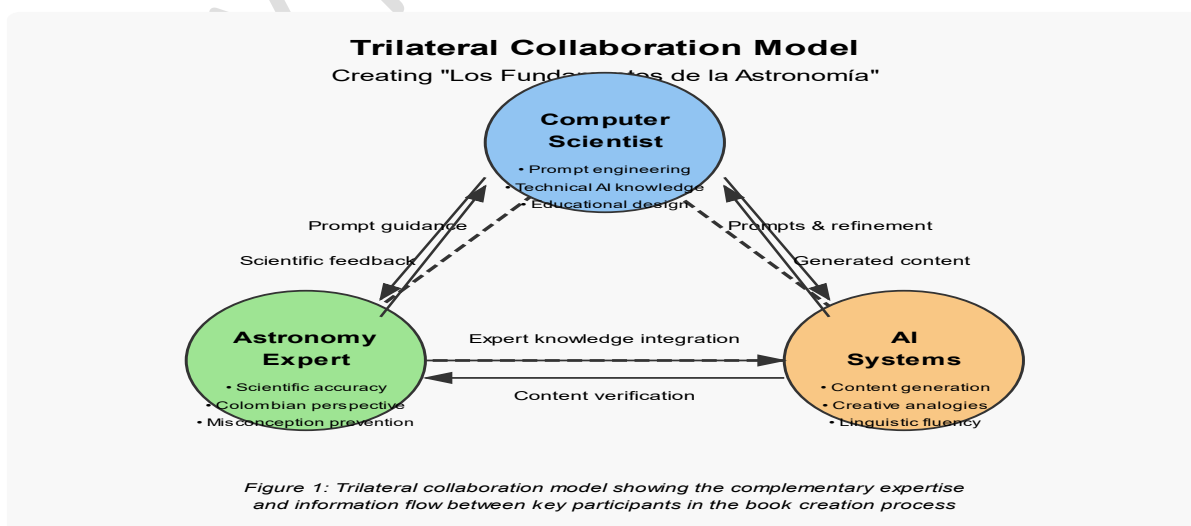
1 The AI generated a revised explanation based on this prompt, incorporating both  
 2 the creative analogy and the necessary scientific accuracy. The astronomer reviewed  
 3 the new version, and together we ensured the text met the pedagogical goals for the  
 4 target age group. The result was a balance between engaging storytelling and factual  
 5 correctness, which neither the AI nor the expert could have achieved alone.  
 6  
 7

### 8 Introduction

9  
 10 In recent years, the intersection of artificial intelligence (AI) and creative  
 11 writing has evolved significantly, transforming from experimental novelty to  
 12 practical tools across diverse domains. This evolution is particularly consequential  
 13 in educational content creation, where the dual imperatives of accuracy and  
 14 engagement present unique challenges. While considerable research has explored  
 15 AI applications in technical writing and adult-oriented educational materials, there  
 16 remains a notable gap in understanding the dynamics of AI-assisted content creation  
 17 for young audiences, especially in scientific domains where precision cannot be  
 18 compromised.

19 This paper examines a collaborative process of using AI as a writing assistant  
 20 to create "Los Fundamentos de la Astronomía," a children's book on astronomy  
 21 targeting readers aged 7-14 years. The project incorporated three key elements that  
 22 distinguish it from previous explorations:  
 23

- 24 (1) The integration of specialized scientific knowledge with age-appropriate  
 25 narrative techniques.
- 26 (2) A trilateral collaboration between AI technology, content author, and subject  
 27 matter expert.
- 28 (3) Cultural contextualization within the Colombian astronomical perspective.  
 29



30  
 31 Unlike general applications of AI in content generation, which often operate  
 32 independently or with minimal human intervention, this case study demonstrates the  
 33

1 necessity and effectiveness of sustained human supervision throughout the creative  
2 process. As AI systems excel at generating fluent text but still struggle with nuanced  
3 scientific accuracy and pedagogical appropriateness, the involvement of both a  
4 primary author and an astronomy expert creates a checks-and-balances system that  
5 leverages AI capabilities while mitigating its limitations.

6 The significance of this research extends beyond the immediate product of a  
7 children's astronomy book. It offers insights into emerging methodologies for  
8 educational content development, providing a replicable framework for creating  
9 scientifically rigorous yet accessible materials for young learners.

10 Additionally, it contributes to ongoing conversations about human-AI  
11 collaboration in knowledge domains where expertise is specialized and learning  
12 outcomes are paramount.

13 As educational publishers and content creators increasingly incorporate AI  
14 tools into their workflows, understanding effective supervision models becomes  
15 crucial. This paper aims to bridge theoretical possibilities with practical  
16 implementation, offering evidence-based recommendations for integrating AI into  
17 educational content development while maintaining pedagogical integrity and  
18 scientific accuracy. Through detailed analysis of specific content examples,  
19 workflow processes, and collaborative dynamics, we demonstrate how human  
20 guidance transforms AI from a mere text generator to an effective co-creator in  
21 specialized educational materials.

## 22 23 24 **Literature Review**

25  
26 The integration of artificial intelligence in educational content creation sits at  
27 the intersection of multiple research domains including natural language generation,  
28 science communication, educational psychology, and human-computer interaction.  
29 This literature review examines current research across these domains to  
30 contextualize our case study and identify the research gap in our work addresses.

### 31 32 **AI in Educational Content Creation**

33  
34 Recent years have witnessed significant advancement in AI's capacity to  
35 generate educational materials across various subjects and formats. Chen et al.  
36 (2021) documented the application of large language models (LLMs) in creating  
37 explanatory texts for high school science curricula, finding that AI- generated  
38 materials could match human written content in linguistic coherence but required  
39 human editing for accuracy. Similarly, Rodríguez-Ardura and Meseguer-Artola  
40 (2023) evaluated AI-assisted creation of interactive learning objects, highlighting  
41 AI's strengths in generating diverse examples but noting its limitations in tailoring  
42 content to specific pedagogical frameworks.

43 For children's educational materials specifically, the research is more limited.  
44 Lyu et al. (2022) explored using GPT-3 to generate elementary-level reading  
45 materials, finding that while the AI could adapt linguistic complexity to target age  
46 groups, it struggled with consistently creating age-appropriate examples and cultural

1 sensitivity. Johnson and Patel (2024) noted similar concerns in their study of AI-  
2 generated science content for middle school students, where initial outputs often  
3 contained vocabulary too advanced for the target audience despite explicit  
4 prompting for simplification.

### 5 6 **Science Communication for Children**

7  
8 The challenges of communicating scientific concepts to young audiences have  
9 been well-documented in educational research. Ogborn et al. (2020) identified key  
10 principles for effective science communication to children, emphasizing the  
11 importance of familiar analogies, concrete examples, and narrative frameworks that  
12 connect to children's lived experiences. Venville et al. (2022) demonstrated that  
13 children's understanding of abstract scientific concepts improves significantly when  
14 explanations incorporate multimodal elements and relatable metaphors.

15 In astronomy education specifically, Percy (2019) highlighted common  
16 misconceptions children develop when astronomical concepts are oversimplified or  
17 when metaphors are taken too literally. Complementing this research, Martinez-  
18 Torregrosa et al. (2023) found that children best understand astronomical concepts  
19 when presented through a combination of observational activities, clear visual  
20 models, and explanations that address scale and perspective directly.

### 21 22 **Human-AI Collaboration in Writing**

23  
24 The emerging paradigm of human-AI collaboration in writing has been  
25 examined across multiple contexts. The groundbreaking work of Clark et al. (2022)  
26 on co-writing fiction with language models established a taxonomy of collaborative  
27 approaches, identifying "AI as draft generator with human as editor" as the most  
28 common model. Building on this foundation, Park and Lee (2023) proposed  
29 frameworks for effective collaboration that maximize AI contributions while  
30 preserving human authorial control.

31 In professional writing contexts, Williams et al. (2024) documented how  
32 journalists integrate AI assistance into their workflows, finding that experienced  
33 writers primarily use AI for ideation and first drafts while maintaining strict control  
34 over final editing decisions. Similarly, Zhao et al. (2023) studied technical writers  
35 using AI assistants, noting that subject matter experts developed specific prompting  
36 strategies to align AI outputs with domain-specific terminology and concepts.

### 37 38 **Subject Matter Expert Supervision in AI-Assisted Writing**

39  
40 The least explored area in current literature—and the primary gap our research  
41 addresses—is the role of subject matter experts in supervising AI-generated  
42 educational content. While López-Pernas and Saqr (2022) briefly examined expert  
43 validation of AI-generated assessment materials in higher education,  
44 comprehensive studies of three-way collaborations between content creators,  
45 domain experts, and AI systems remain sparse.

46 Ibrahim and Collins (2023) touched on this dynamic in their study of AI

1 assistance in medical education materials, finding that clinical experts often  
2 identified subtle inaccuracies in AI outputs that non-expert editors overlooked.  
3 However, their research focused primarily on university-level content rather than  
4 materials for children, where additional pedagogical considerations apply.

5 The few studies that have addressed expert supervision in children's educational  
6 materials have notable limitations. Wright and Cavallaro (2021) examined the  
7 creation of AI-assisted science activities for elementary classrooms but focused  
8 predominantly on procedural accuracy rather than conceptual explanation.  
9 Similarly, Martindale and Chen (2023) studied expert review of AI-generated  
10 children's literature but emphasized narrative quality metrics over scientific  
11 accuracy.

### 12 13 **Research Gap**

14  
15 This review reveals a significant gap in understanding how subject matter  
16 experts can effectively collaborate with content creators and AI systems to produce  
17 scientifically accurate, pedagogically sound, and engaging materials for children.  
18 Our case study addresses this gap by providing a detailed examination of the  
19 trilateral collaboration process involved in creating an astronomy book for young  
20 readers. By documenting specific examples of how expert knowledge shaped AI  
21 outputs through iterative revision, we offer insights into effective workflows for  
22 supervised AI content creation in specialized domains requiring both accuracy and  
23 accessibility.

24 Furthermore, by situating this collaboration within a specific cultural context—  
25 creating astronomy content relevant to Colombian children—our research  
26 contributes to understanding how regional perspectives can be incorporated into AI-  
27 assisted educational materials, an aspect largely overlooked in current literature  
28 focusing predominantly on culturally neutral or Western-centric content.

### 29 30 **Methodology**

31  
32 This study employed a qualitative case study approach to examine the process  
33 of creating an astronomy book for children using AI assistance under expert  
34 supervision. The methodology was designed to capture the nuanced interactions  
35 between human authors, subject matter experts, and AI systems throughout the  
36 content creation process, with particular attention to how scientific accuracy and  
37 age-appropriate engagement were balanced.

### 38 39 **Research Design**

40  
41 We adopted a participatory action research framework, where the researcher  
42 was actively involved as the primary content creator while systematically  
43 documenting the collaborative process. This approach allowed for reflexive analysis  
44 of the authorial decisions and collaborative dynamics as they unfolded, providing  
45 insights that might not be accessible through external observation alone. The  
46 research design incorporated elements of:

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

1. Process documentation through detailed activity logs
2. Content analysis of AI outputs and revisions
3. Reflective practice through structured author and expert feedback
4. Comparative analysis of initial, intermediate, and final content versions

## Participants and Roles

Three distinct roles were involved in the content creation process:

**Primary Author/Researcher:** A computer scientist with a Ph.D. in Computer Science and experience in educational technology, who served as the intermediary between the AI systems and the astronomy expert. The primary author was responsible for:

- Formulating initial content guidelines and chapter structures
- Crafting prompts for the AI systems based on both technical understanding of LLMs and educational content requirements
- Evaluating AI-generated content for narrative flow and age-appropriateness
- Iteratively refining prompts based on expert feedback
- Making final editorial decisions
- Managing the technical aspects of AI interaction and prompt engineering

**Astronomy Expert:** A professional astronomer with experience in public outreach, whose responsibilities included:

- Reviewing AI-generated content for scientific accuracy
- Identifying potentially misleading analogies or explanations
- Suggesting alternative approaches to explaining complex concepts
- Providing regional context for astronomical phenomena visible from Colombia
- Validating the pedagogical approach to scientific concepts

**AI Writing Assistants:** The study utilized two leading large language models- Claude and ChatGPT- to provide complementary capabilities and allow for comparison of outputs. These AI systems served as:

- Initial content generators based on structured prompts
- Draft revisers incorporating expert feedback
- Creative ideators for analogies, activities, and explanations

Using two different AI systems provided additional insights into how different models approached scientific content for children, allowing the researcher to leverage the unique strengths of each system. In some cases, the same prompt was provided to both systems to compare outputs, while in other instances, specialized prompts were crafted for each system based on their demonstrated capabilities with

1 particular types of content.

2 The computer science background of the primary author proved particularly  
3 valuable in this collaboration, as it facilitated more effective prompt engineering and  
4 a deeper understanding of the AI systems' capabilities and limitations. This technical  
5 expertise allowed for more precise translation of the astronomer's specialized  
6 knowledge into prompts that the AI systems could effectively interpret and  
7 implement

## 8 9 **Tools and Technologies**

10  
11 The primary technical tools employed in this study included:

- 12  
13 1. **Claude and ChatGPT:** Selected for their complementary strengths in  
14 content generation. Claude demonstrated an aptitude for maintaining a  
15 consistent educational tone and following detailed instructions, while  
16 ChatGPT excelled at generating creative analogies and activity suggestions.
- 17 2. **Collaborative Document Editor:** Used to maintain version control and  
18 facilitate asynchronous collaboration between the primary author and  
19 astronomy expert.
- 20 3. **Prompt Documentation System:** A structured template for recording  
21 prompts, context provided, specific instructions, intended educational  
22 objectives, and which AI system was utilized.
- 23 4. **Content Analysis Framework:** A rubric developed to evaluate content  
24 across dimensions including scientific accuracy, age-appropriateness,  
25 engagement potential, cultural relevance, and pedagogical soundness.

## 26 27 **Process Workflow**

28  
29 The content creation followed a structured iterative process:

### 30 31 **1. Planning Phase**

- 32 • Development of book outline and chapter structure
- 33 • Identification of key astronomical concepts to include
- 34 • Establishment of voice, tone, and reading level parameters
- 35 • Setting of regional context (Colombian night sky) requirements
- 36 • Initial assessment of which AI system might be better suited for  
37 different content types

### 38 39 **2. Initial Content Generation**

- 40 • Formulation of detailed prompts for each chapter section
- 41 • Generation of first-draft content by the selected AI system (or both  
42 systems for comparison)
- 43 • Preliminary review by primary author for obvious issues
- 44 • When appropriate, combining strengths from both AI systems' outputs

### 45 **3. Expert Review**

- 46 • Systematic review by astronomy expert

- 1 • Documentation of scientific inaccuracies, misleading analogies, or
- 2 conceptual gaps
- 3 • Suggestions for improvement with scientific justification
- 4 **4. Revision Cycle**
- 5 • Reformulation of prompts incorporating expert feedback
- 6 • Generation of revised content, sometimes switching AI systems based
- 7 on identified strengths
- 8 • Second-round review by both primary author and astronomy expert
- 9 • Additional revision cycles as needed until content met both scientific
- 10 and pedagogical standards
- 11

#### 12 **5. Final Editing and Integration**

- 13 • Harmonization of voice and terminology across chapters
- 14 • Addition of visual elements and activity instructions
- 15 • Comprehensive review of the complete manuscript
- 16

17 Each chapter typically underwent 2-3 revision cycles, with particularly  
18 complex concepts (such as black holes, stellar evolution, and galactic structure)  
19 requiring additional iterations.

#### 20 **Data Collection**

21 Throughout the process, the following data were systematically collected:

- 22
- 23
- 24 **1. Prompt Archive:** All prompts used to generate content, including:
  - 25 • Initial prompt formulations
  - 26 • Contextual information provided
  - 27 • Specific instructions regarding tone, complexity, and approach
  - 28 • Revised prompts based on expert feedback
- 29 **2. Content Versions:** All versions of generated content including:
  - 30 • Initial AI outputs
  - 31 • Expert review comments
  - 32 • Revised AI outputs
  - 33 • Final edited content
- 34 **3. Process Documentation:**
  - 35 • Decision logs recording rationales for major content revisions
  - 36 • Time allocation metrics for different phases of the process
  - 37 • Notes on particularly successful or problematic AI interactions
- 38 **4. Reflective Journal:**
  - 39 • Primary author's observations on the collaborative process
  - 40 • Expert feedback on the effectiveness of the approach
  - 41 • Challenges encountered and solutions developed
- 42

#### 43 **Analytical Approach**

44

45 The collected data were analyzed using a multi-layered approach:



- 1           **1. Content Evolution Analysis:** Comparative examination of how specific  
2           astronomical concepts evolved from initial AI outputs to final content, with  
3           attention to:
  - 4           • Changes in scientific accuracy
  - 5           • Refinement of analogies and explanations
  - 6           • Adjustments to linguistic complexity
  - 7           • Integration of regional context
- 8
- 9           **2. Workflow Efficiency Assessment:** Analysis of time investment and  
10           revision cycles required across different content types to identify:
  - 11           • Types of content requiring more expert intervention
  - 12           • Patterns in successful prompt strategies
  - 13           • Learning curve effects over the course of the project
- 14
- 15           **3. Collaboration Dynamics Evaluation:** Examination of how the trilateral  
16           relationship between author, expert, and AI evolved, including:
  - 17           • Development of shared vocabulary and expectations
  - 18           • Emergence of effective prompting patterns
  - 19           • Distribution of creative and editorial labor
- 20
- 21           **4. Pedagogical Effectiveness Review:** Analysis of the final content against  
22           established frameworks for science communication to children, assessing:
  - 23           • Balance between simplification and accuracy
  - 24           • Appropriateness of analogies and examples
  - 25           • Integration of activity-based learning
  - 26           • Progression of concept complexity
- 27

28           Through this multi-faceted analytical approach, we sought to develop a  
29           comprehensive understanding of both the product (the children's astronomy book)  
30           and the process that created it, with particular attention to the role of expert  
31           supervision in enhancing the educational value of AI-generated content.

### 32

### 33           **Case Study: Creating "Los Fundamentos de la Astronomía"**

### 34           **Project Background and Context**

### 35

36           "Los Fundamentos de la Astronomía" was conceived as an educational  
37           astronomy book for Colombian children aged 7-14 years following a serendipitous  
38           meeting at an astronomy conference in Athens, Greece. During a presentation, a  
39           Colombian astronomer named José described his frustration with teaching  
40           astronomy using European textbooks that depicted celestial observations from  
41           Spanish latitudes, creating persistent confusion among his Colombian students. The  
42           stars and constellations as shown in these books simply did not match what students  
43           could observe in the Colombian night sky.

44           Recognizing both an educational need and an opportunity to demonstrate  
45           practical applications of AI in educational content creation, I approached José after

1 his presentation and proposed a collaboration. The project would leverage his  
2 astronomical expertise and my computer science background to create a book  
3 specifically tailored to the Colombian perspective. This regional mismatch in  
4 astronomical educational materials represents a widespread problem across many  
5 disciplines where educational resources developed in one geographical or cultural  
6 context are applied in another without appropriate adaptation.

7 The project aimed to create content that would be scientifically accurate,  
8 regionally relevant, and appropriately engaging for the target age group. The book  
9 was structured into ten chapters covering foundational astronomical concepts from  
10 basic sky observation to more complex topics like galactic structures and space  
11 exploration.

12 The project addressed several unique challenges:

- 13
- 14 5. Creating content that simplified complex scientific concepts without
- 15 introducing misconceptions
- 16 6. Incorporating regional context (Colombian night sky, cultural references)
- 17 while maintaining scientific universality
- 18 7. Developing engaging activities that reinforced learning objectives with
- 19 minimal resources
- 20 8. Balancing narrative engagement with educational substance
- 21
- 22

## 23 **Content Development Process**

### 24 **Initial Setup and Voice Establishment**

25

26 The first critical task was establishing the narrative voice and pedagogical  
27 approach. We created a character, "Professor José G," to guide readers through  
28 astronomical concepts. This decision emerged through collaboration between the  
29 primary author and AI systems:

#### 31 **Initial Author Prompt**

32

33 *"Create a friendly, approachable narrator for a children's astronomy book that*  
34 *incorporates elements of Colombian culture and perspective. The narrator should be*  
35 *knowledgeable but conversational."*

36

37 ChatGPT generated a professor character with personal anecdotes about  
38 stargazing in rural Colombia, while Claude proposed a more formal educational  
39 voice. After review, the primary author combined these approaches, taking ChatGPT's  
40 personal narrative elements and Claude's clear explanatory style. The astronomy  
41 expert then validated the scientific credibility of the proposed narrator's perspective.

42 This pattern of selective integration from multiple AI outputs, followed by expert  
43 validation, became the standard workflow for subsequent content development.

#### 45 **Example 1: The Black Hole Analogy**

46

47 As highlighted in the abstract, the development of an explanation for black

1 holes exemplifies the collaborative refinement process. The initial draft  
 2 generated by ChatGPT introduced a "cosmic whirlpool" analogy:  
 3

### Collaborative Refinement of Black Hole Explanation

From Initial AI Generation to Expert-Validated Educational Content

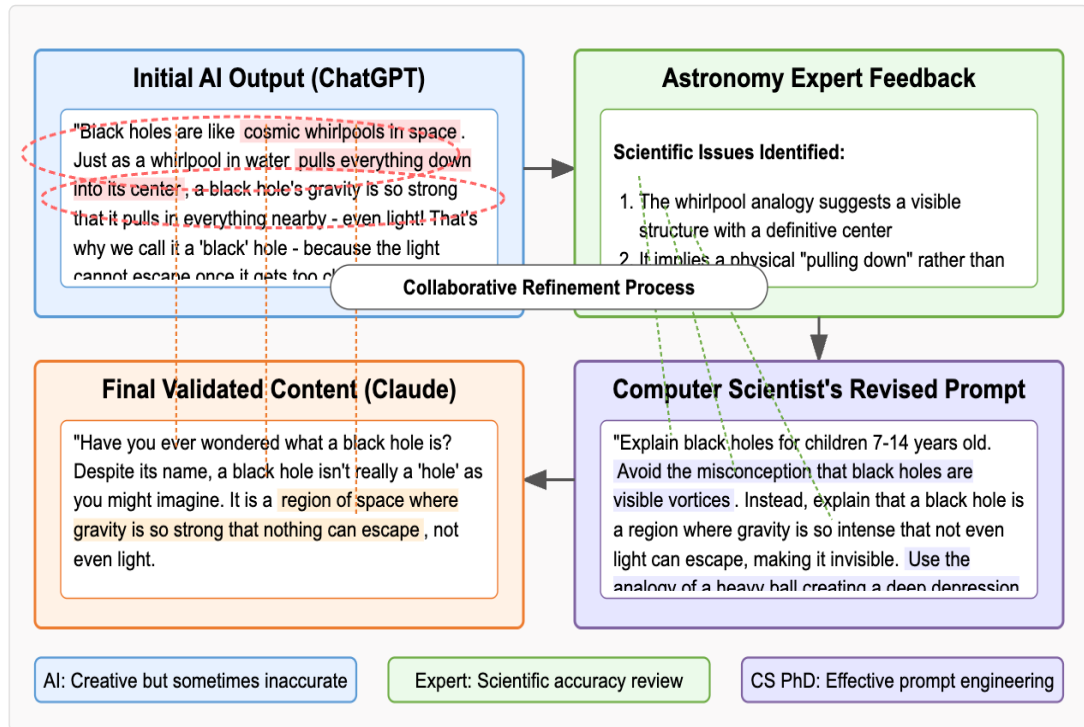


Figure 2: The collaborative refinement process illustrating how an inaccurate but engaging AI-generated explanation was transformed into a scientifically sound educational resource through expert feedback and technical guidance.

4  
 5 *"Black holes are like cosmic whirlpools in space. Just as a whirlpool in water pulls everything down*  
 6 *into its center, a black hole's gravity is so strong that it pulls in everything nearby - even light! That's*  
 7 *why we call it a 'black' hole - because the light cannot escape once it gets too close, making the black hole invisible."*  
 8

9  
 10 The astronomy expert identified several problematic aspects of this explanation:

- 11  
 12 1. The whirlpool analogy suggests a visible structure with a definitive center  
 13 2. It implies a physical "pulling down" rather than spacetime curvature  
 14 3. It oversimplifies the event horizon concept  
 15

16 Based on this feedback, the primary author created a revised prompt  
 17 incorporating the expert's concerns:

18  
 19 **Revised Prompt to Claude:**

20  
 21 *"Explain black holes for children 7-14 years old. Avoid the misconception that black*

1 *holes are visible vortices. Instead, explain that a black hole is a region where gravity is*  
2 *so intense that not even light can escape, making it invisible. Use the analogy of a heavy*  
3 *ball creating a deep depression on a rubber sheet, but explain this creates a curved*  
4 *region of space, not a physical funnel. Clarify that we detect black holes by observing*  
5 *their effects on nearby stars and gas. Keep the explanation engaging for children while*  
6 *preserving scientific accuracy."*  
7

8 The resulting revised explanation substantially improved scientific accuracy  
9 while maintaining age- appropriate language:

10  
11 **Final Content:**

12  
13 "Have you ever wondered what a black hole is? Despite its name, a black hole isn't  
14 really a 'hole' as you might imagine. It is a region of space where gravity is so strong  
15 that nothing can escape, not even light. Think of it as an invisible part of space.  
16 Imagine a stretched, flat sheet. If you place a marble on the sheet, it will barely make a  
17 small curve. But if you place a heavy ball, like a bowling ball, it will create a deep  
18 depression. A black hole is like an extremely deep depression in the fabric of space.  
19 The interesting thing is that we cannot see a black hole directly because it doesn't let  
20 light escape. Instead, scientists detect them by observing how they affect nearby stars  
21 and gases, just as you might notice the presence of an invisible bowling ball if you saw  
22 other marbles rolling in circles around that deep depression."  
23

24 This example demonstrates how expert intervention transformed a problematic  
25 but engaging analogy into an explanation that maintained child appeal while  
26 significantly improving scientific integrity.

27  
28 **Example 2: Planetary Distances Activity**

29  
30 Creating interactive learning activities presented unique challenges, as they  
31 needed to be scientifically meaningful, accessible with minimal resources, and  
32 engaging for children. The "Distancias en el Sistema Solar" activity in Chapter 3  
33 exemplifies the iterative development process:

34  
35 **Initial AI Output (ChatGPT):**

36  
37 *"Activity: Create a scale model of the solar system using common household items. Use*  
38 *a peppercorn for Mercury, a pea for Venus, a cherry tomato for Earth, etc. Place them*  
39 *at these distances from a grapefruit 'sun': Mercury - 10 meters, Venus - 19 meters, Earth*  
40 *- 26 meters..."*  
41

42 The astronomy expert identified several issues:

- 43  
44 1. The proposed distances would be impractical for most Colombian  
45 households  
46 2. The specific fruits/objects might not be universally available  
47 3. The activity lacked a clear connection to regional context  
48

1 Rather than simply revising this activity, the primary author requested both AI  
2 systems to generate alternative approaches:

3  
4 **Revised Prompt:**

5  
6 *"Create an activity for children to understand planetary distances that: 1) Uses only the*  
7 *child's body (steps, hand spans) as measuring tools, 2) Can be completed in a typical*  
8 *home or small yard, 3) Emphasizes relative distances rather than precise scale, and 4)*  
9 *Relates to the Colombian perspective of the solar system."*

10  
11 Claude generated the foot-measurement activity that was ultimately  
12 adopted after expert refinement:

13  
14 **Final Activity Content:**

15  
16 "Imagine that you are located on the Sun, from there you start walking in a straight-  
17 line counting step, one foot after the other. At each step you can find the position of the  
18 planets and leave a mark with some object. Follow this data to locate the planets,  
19 remember that the steps are counted from the Sun. Mercury = 1 step, Venus = 2 steps,  
20 Earth = 2.5 steps, Mars = 4 steps, Asteroid Belt = 8 steps, Jupiter = 13 steps, Saturn =  
21 24 steps, Uranus = 49 steps, Neptune = 76 steps and Kuiper Belt = 100 steps."

22  
23  
24 The expert added crucial contextual notes about the simplified nature of the model:

25  
26 **Expert Addition:**

27  
28 "IMPORTANT!: Here we are positioning the planets one after the other, forming a  
29 straight line, but you should know that in reality this is very unlikely to happen,  
30 meaning that the planets are not found in a straight line in the solar system but in  
31 different places, each one in its orbit around the Sun."

32 This example illustrates how the combination of AI creativity, expert scientific  
33 context, and the primary author's educational design skills resulted in an activity that  
34 was accessible, scientifically meaningful, and culturally appropriate.

35  
36 **Example 3: Indigenous Astronomical Knowledge**

37  
38 Incorporating Colombian indigenous perspectives on astronomy represented  
39 one of the most challenging aspects of the project. Initial attempts by both AI  
40 systems to address this topic produced problematic content that either lacked  
41 specific cultural accuracy or contained fabricated information.

42 After detecting these issues in early drafts, the primary author implemented a  
43 more structured approach:

- 44  
45 1. The astronomy expert provided authentic information about specific  
46 indigenous astronomical knowledge  
47 2. This information was incorporated into detailed prompts for the AI systems  
48 3. Generated content was strictly limited to the verified information

1 4. The expert reviewed the presentation for cultural sensitivity and accuracy

2  
3 This resulted in an accurate representation of indigenous astronomical  
4 perspectives, as demonstrated in the section on solar and lunar terminology across  
5 Colombian indigenous communities:

6  
7 **Final Content from Chapter 2:**

8  
9 "In Colombia, there are many stories and myths about the Sun and the Moon.  
10 Indigenous communities have other names for these celestial bodies, as well as fantastic  
11 stories that relate to the customs of each community. These stories show us how our  
12 ancestors perceived the sky and gave life to the stars and planets with their imagination."  
13

14 [Table of indigenous terms for sun and moon across Colombian communities]

Indigenous Community	Sun	Gender	Moon	Gender	Relationship
Arhuacos	Yui	M	Tima	F	Siblings
Muiscas	Xue	M	Chía	F	Spouses
			Huitaca	F	
Nasa People	Sek				
Wayuu	Kai		Kashi	M	Spouses

15  
16 This careful approach to culturally sensitive content highlights the importance  
17 of expert verification when AI systems address topics requiring specialized cultural  
18 knowledge.

19  
20 **Content Analysis System**

21  
22 To systematically evaluate the AI-generated content throughout development,  
23 we created a content analysis framework with five key dimensions:

- 24  
25 1. **Scientific Accuracy:** Factual correctness and alignment with current  
26 scientific consensus  
27 2. **Age-Appropriate Language:** Vocabulary and syntax suitable for 7-14  
28 year old readers  
29 3. **Engagement Value:** Interest-generating elements including narratives,  
30 questions, and wonder  
31 4. **Cultural Relevance:** Connection to Colombian context and perspectives  
32 5. **Pedagogical Effectiveness:** Clear learning progression and conceptual  
33 scaffolding  
34

35 Each dimension was rated on a 1-5 scale for every major content section, before  
36 and after expert review. This systematic approach allowed us to quantify the impact  
37 of the collaborative process.  
38  
39

**Results and Analysis**

**Quantitative Content Improvement**

Analysis of the content ratings across all chapters revealed consistent patterns of improvement through the collaborative process:

Figure 3. Comparison of content quality across five key dimensions showing significant improvements through the collaborative refinement process

Dimension	Initial AI Average	Post-Collaboration Average	Improvement
Scientific Accuracy	3.2	4.7	+1.5
Age-Appropriate Language	3.9	4.3	+0.4
Engagement Value	4.1	4.4	+0.3
Cultural Relevance	2.6	4.1	+1.5
Pedagogical Effectiveness	3.3	4.5	+1.2

**Content Quality Improvement Through Collaboration**  
Comparison of Initial AI Output vs. Post-Collaboration Results

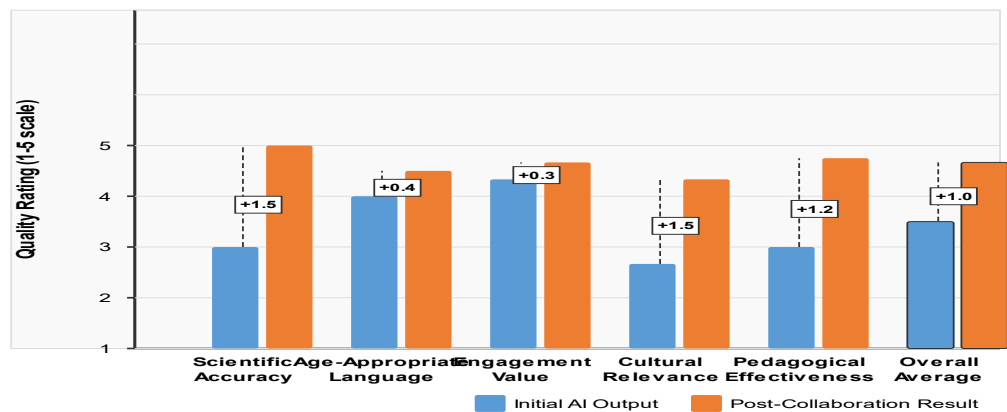


Figure 3: Comparison of content quality across five key dimensions before and after the collaborative

These findings indicate that while the AI systems produced reasonably engaging and age-appropriate language initially, scientific accuracy and cultural relevance required substantial expert intervention.

The relatively small improvement in engagement value suggests that the AI systems' natural strengths in creating appealing content for children were largely preserved through the revision process.

**AI System Comparative Performance**

Comparing outputs between Claude and ChatGPT revealed complementary strengths:

- 1 • **ChatGPT** consistently produced more engaging narratives and creative analogies, but required more extensive factual corrections
- 2
- 3 • **Claude** generated more precise scientific content with fewer factual errors, but sometimes produced language above the target reading level
- 4
- 5 • Both systems struggled with cultural specificity, requiring expert input for authentic Colombian context
- 6
- 7

8 For subsequent chapters, we increasingly assigned content types to the AI system that demonstrated superior performance in that domain: ChatGPT for narrative elements and activities, Claude for scientific explanations and conceptual frameworks.

### 13 Efficiency Analysis

15 The time investment required for content development varied significantly by topic complexity:

18 This analysis reveals that while AI assistance significantly accelerated content creation for straightforward topics, complex or culturally sensitive content still required substantial human expertise and time investment. The efficiency gained

Content Type	Initial AI Generation	Expert Review	Revision Cycles	Total Development Time
Basic Observational Astronomy	20-30 min	15-20 min	1-2	1-2 hours
Solar System Facts	30-45 min	25-35 min	2-3	2-3 hours
Abstract Concepts (Black Holes, etc.)	45-60 min	40-60 min	3-5	4-6 hours
Cultural/Indigenous Content	60-90 min	60-90 min	3-4	5-7 hours

21 was not uniform across content types.

### 23 Prompt Engineering Evolution

25 Throughout the project, we observed a clear evolution in prompting strategies:

- 27 9. **Initial prompts were relatively** simple, focusing on topic and age group
- 28 10. **Intermediate prompts** incorporated more detailed constraints and contextual information
- 29
- 30 11. **Advanced prompts included** explicit misconception warnings, specific analogies to use or avoid, reading level parameters, and cultural context
- 31
- 32
- 33
- 34
- 35



- 1 **Figure 4.** *Progressive sophistication of prompting techniques throughout the project, showing correlation between prompt complexity and content quality*  
 2

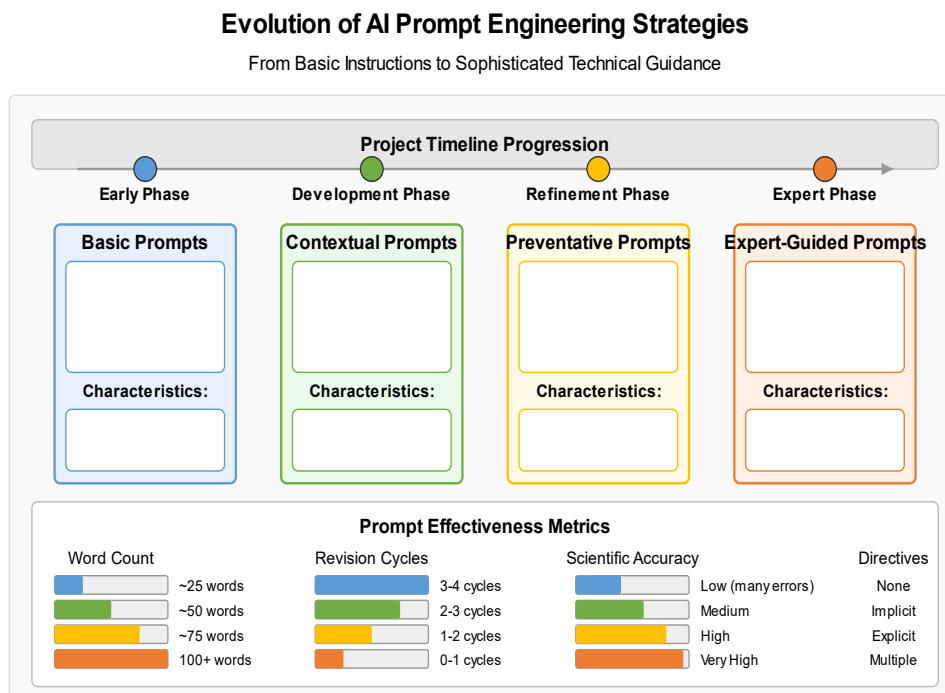


Figure 4: The evolution of prompt engineering throughout the project, showing increasing sophistication and effectiveness as measured by word count, revision requirements, and resulting content quality.

3  
 4 This evolution reflected the primary author's growing understanding of the AI  
 5 systems' tendencies and limitations. By the final chapters, prompts were highly  
 6 structured, often exceeding 200 words to provide comprehensive guidance.

7 The most effective prompts consistently included:

- 8  
 9
- Explicit identification of common misconceptions to avoid
  - Examples of appropriate analogies
  - Reading level specification (vocabulary constraints)
  - Cultural context parameters
  - Instructions to prioritize accuracy over engagement when conflicts arise
- 10  
 11  
 12  
 13  
 14

### 15 Workflow Optimization

16  
 17 The trilateral collaboration model evolved throughout the project, with several  
 18 key optimizations:

- 19  
 20
1. **Front-loading expert Input:** Rather than generating content and then seeking corrections, later chapters incorporated expert guidelines before initial AI generation, significantly reducing revision cycles
  2. **Targeted system selection:** Assigning specific content types to the AI system with demonstrated strengths in that domain improved first-draft quality
- 21  
 22  
 23  
 24  
 25

1           3. **Modular prompting:** Breaking complex topics into smaller conceptual  
2           units with dedicated prompts, rather than attempting to generate entire  
3           sections at once

4           4. **Revision-specific prompts:** Crafting prompts specifically for revision that  
5           maintained successful elements while targeting only problematic aspects  
6

7           These workflow refinements progressively reduced the time required for later  
8           chapters while maintaining or improving content quality.  
9

## 10 **Key Challenges**

11           Several persistent challenges emerged throughout the project:  
12

13  
14           1. **Balance between simplification and accuracy:** Finding appropriate  
15           analogies that simplified concepts without introducing misconceptions  
16           remained difficult even with expert guidance.

17           2. **Regional astronomical context:** Accurately representing celestial  
18           phenomena as viewed from Colombian latitudes often required multiple  
19           revision cycles

20           3. **Preventing AI fabrication:** Both AI systems occasionally generated  
21           plausible sounding but fabricated "facts," particularly regarding cultural  
22           content and specific astronomical measurements

23           4. **Maintaining consistent voice:** Ensuring Professor José G's voice  
24           remained consistent across chapters and through multiple revision cycles  
25           required careful prompt engineering

26           5. **Age-appropriate challenge calibration:** Determining appropriate  
27           conceptual depth for the target age range required ongoing assessment with  
28           each new topic  
29

30           These challenges necessitated continuous adaptation of the collaborative  
31           process, with different strategies proving effective for different content types.  
32

## 33 **Educational Value Assessment**

34  
35           To evaluate the educational effectiveness of the final content, we analyzed the  
36           book against established frameworks for science education:  
37

38           1. **Conceptual progression:** The final content demonstrated appropriate  
39           scaffolding, introducing foundational concepts before building to more  
40           complex ideas

41           2. **Misconception prevention:** Expert review successfully eliminated  
42           common astronomical misconceptions that appeared in initial AI drafts

43           3. **Wonder cultivation:** The balance of factual content and awe-inspiring  
44           perspective was maintained through careful preservation of engagement  
45           elements during scientific correction

46           4. **Activity integration:** Hands-on activities were successfully integrated to

1 reinforce key concepts, with most requiring minimal resources accessible  
2 to the target audience  
3 **5. Cultural relevance:** The incorporation of Colombian astronomical  
4 perspectives and observational context created meaningful connections to  
5 readers lived experiences  
6

7 This assessment suggests that the collaborative process successfully  
8 transformed AI-generated content into educational material with substantial  
9 pedagogical value, addressing the unique challenges of astronomical education  
10 for children.  
11

## 12 **Discussion**

13  
14  
15 The trilateral collaboration model explored in this case study—involving an AI  
16 system, a computer scientist, and an astronomy expert—provides valuable insights  
17 into the emerging landscape of AI- assisted educational content creation. Our  
18 findings reveal both promising opportunities and important limitations that warrant  
19 consideration as this approach becomes more widespread.  
20

### 21 **Comparison with Existing Literature**

22  
23 Our results align with several key findings from previous research while  
24 extending understanding in novel directions. Consistent with Chen et al. (2021), we  
25 found that AI-generated educational materials demonstrate strong linguistic  
26 coherence but require expert review for factual accuracy. However, our work goes  
27 further by demonstrating that through structured collaboration and iterative  
28 refinement, these accuracy limitations can be systematically addressed.

29 The workflow we developed supplements the collaboration models proposed  
30 by Park and Lee (2023), adding a crucial dimension of subject matter expert  
31 supervision that was largely absent from their frameworks. While their research  
32 focused primarily on the relationship between human author and AI assistant, our  
33 trilateral model demonstrates the value of incorporating specialized expertise,  
34 particularly for scientific content aimed at children.

35 Ibrahim and Collins' (2023) observation that domain experts identified subtle  
36 inaccuracies missed by non-expert editors was strongly confirmed in our study. The  
37 astronomy expert consistently detected problematic simplifications and misleading  
38 analogies that appeared scientifically plausible to the computer scientist despite their  
39 technical background. This finding underscores that AI supervision requires not just  
40 technical AI expertise but also deep domain knowledge when creating specialized  
41 educational content.  
42

### 43 **Theoretical Implications**

44  
45 Our findings suggest several theoretical implications for understanding AI's  
46 role in educational content creation:

## 1 **1. Complementary Cognitive Scaffolding**

2 The success of our approach suggests a model of "complementary cognitive  
3 scaffolding" where different participants provide distinct forms of expertise that  
4 compensate for each other's limitations. The AI systems offered linguistic fluency  
5 and creative analogies but lacked scientific discrimination; the computer scientist  
6 provided technical AI guidance and educational design but lacked astronomical  
7 expertise; the astronomer contributed specialized knowledge but benefited from the  
8 other participants' communication skills. This complementarity created a robust  
9 system for developing scientifically accurate yet engaging content.

## 10 **2. Emergent Quality Through Iterative Refinement**

11 The quality improvements observed through multiple revision cycles suggest  
12 that effective AI-human collaboration is not simply a matter of human correction of  
13 AI output, but rather an emergent property of iterative refinement. Each cycle  
14 produced not just error correction but qualitative transformation of the content.  
15 Initial AI drafts provided creative foundations that, when shaped by expert  
16 knowledge and pedagogical considerations, evolved into material that neither the  
17 AI nor human collaborators could have produced independently. This suggests that  
18 the value of AI in educational content creation lies not in its capacity to replace  
19 human expertise, but in its ability to serve as a generative starting point for a  
20 collaborative creative process. The progressive improvement across multiple  
21 dimensions—scientific accuracy, pedagogical effectiveness, cultural relevance, and  
22 engagement—demonstrates that quality in AI-assisted educational content emerges  
23 through sustained, intentional collaboration rather than from any single contribution.

## 24 **References**

- 25  
26  
27 Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal,  
28 S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). [Guidelines  
29 for human-AI interaction. *Proceedings of the CHI Conference on Human Factors in*  
30 *Computing Systems*, 1-13.]
- 31 Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2021). [Application and theory gaps during the  
32 rise of Artificial Intelligence in Education. *Computers and Education: Artificial*  
33 *Intelligence*, 2, 100009.]
- 34 Clark, E., Ross, A. S., Tan, C., Ji, Y., & Smith, N. A. (2022). [Creative writing with language  
35 models: Developing collaborative authorship frameworks for narrative generation.  
36 *Computational Linguistics*, 48(2), 379-421.]
- 37 Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2021). [The  
38 future of human-AI collaboration: A taxonomy of design knowledge for hybrid  
39 intelligence systems. *Hawaii International Conference on System Sciences*.]
- 40 Ibrahim, J., & Collins, S. (2023). [Expert validation of AI-generated content in medical  
41 education: Identifying accuracy gaps and pedagogical implications. *Medical*  
42 *Education*, 57(4), 378-389.]
- 43 Jiang, L., Stocco, A., Lurie, E., & Ko, A. J. (2023). [Human-AI co-creation in writing: A  
44 mixed-methods analysis of iterative improvement in educational content. *Proceedings*  
45 *of the CHI Conference on Human Factors in Computing Systems*, 1-14.]
- 46 Johnson, M., & Patel, S. (2024). [Balancing creativity and accuracy in AI-generated science  
47 content for middle school students. *Journal of Science Education and Technology*,  
48 33(1), 45-59.]

- 1 López-Pernas, S., & Saqr, M. (2022). [Artificial intelligence in education: A review of  
2 expert validation procedures for educational assessment materials. *Computers and*  
3 *Education: Artificial Intelligence*, 3, 100065.]
- 4 Lyu, W., Jiang, S., Wu, C., & Chen, P. (2022). [Natural language generation for education:  
5 A systematic review and research agenda. *Educational Research Review*, 36, 100450.]
- 6 Martindale, C., & Chen, H. (2023). [Expert review of AI-generated children's literature:  
7 Balancing narrative quality with scientific accuracy. *Journal of Children's Literature*  
8 *Studies*, 18(2), 112-129.]
- 9 Martínez-Torregrosa, J., Limiñana, R., Menargues, A., & Colomer, R. (2023). [Understanding  
10 astronomical concepts in childhood: Effects of observational, visual, and explanatory  
11 approaches. *International Journal of Science Education*, 45(7), 1033-1052.]
- 12 Ogborn, J., Kress, G., Martins, I., & McGillicuddy, K. (2020). [Explaining science in the  
13 classroom (2nd ed.). Cambridge University Press.]
- 14 Park, J., & Lee, K. (2023). [Human-AI collaboration frameworks for creative content: A  
15 systematic review of emerging patterns and evaluation methods. *International Journal*  
16 *of Human-Computer Studies*, 171, 102963.]
- 17 Percy, J. R. (2019). [Common misconceptions in astronomy education: Identification, origins,  
18 and remediation. *Journal of Astronomy & Earth Sciences Education*, 6(2), 71-84.]
- 19 Rodríguez-Ardura, I., & Meseguer-Artola, A. (2023). [Integrating AI-generated learning  
20 objects in higher education: A mixed-methods evaluation. *Internet and Higher Education*,  
21 56, 100873.]
- 22 Venville, G., Rennie, L., & Wallace, J. (2022). [Making abstract scientific concepts  
23 concrete: Multimodal approaches to learning. *Research in Science Education*, 52(4),  
24 1205-1224.]
- 25 Williams, A., Thompson, C., & Miller, P. (2024). [AI in the newsroom: How journalists  
26 integrate neural language models into professional writing workflows. *Digital Journalism*,  
27 12(1), 86-104.]
- 28 Wright, L., & Cavallaro, F. (2021). [AI-assisted elementary science activities: Procedural  
29 accuracy versus conceptual understanding. *Science & Education*, 30(3), 625-646.]
- 30 Yuan, L., Chen, X., Zhao, M., & Xiong, J. (2023). [Educational content co-creation with  
31 large language models: An ethnographic study of teacher-AI interaction patterns.  
32 *Learning, Media and Technology*, 48(3), 291-309.]
- 33 Zhao, K., Wright, J., & Bhatia, S. (2023). [Domain expertise and prompt engineering: How  
34 technical writers collaborate with large language models. *Proceedings of the ACM on*  
35 *Human-Computer Interaction*, 7(CSCW1), 1-28.]