

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

*By Elizabeth D. Diaz**

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 Astronomia**" 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.

Keywords: 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

*Associate Professor, University of Texas at Arlington, USA.

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."

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

Introduction

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

This project also raised important cultural and ethical considerations, particularly in avoiding bias, respecting indigenous knowledge, and ensuring that AI-generated content aligns with pedagogical and developmental needs of young learners. The collaborative framework was designed to address these concerns through continuous expert oversight.

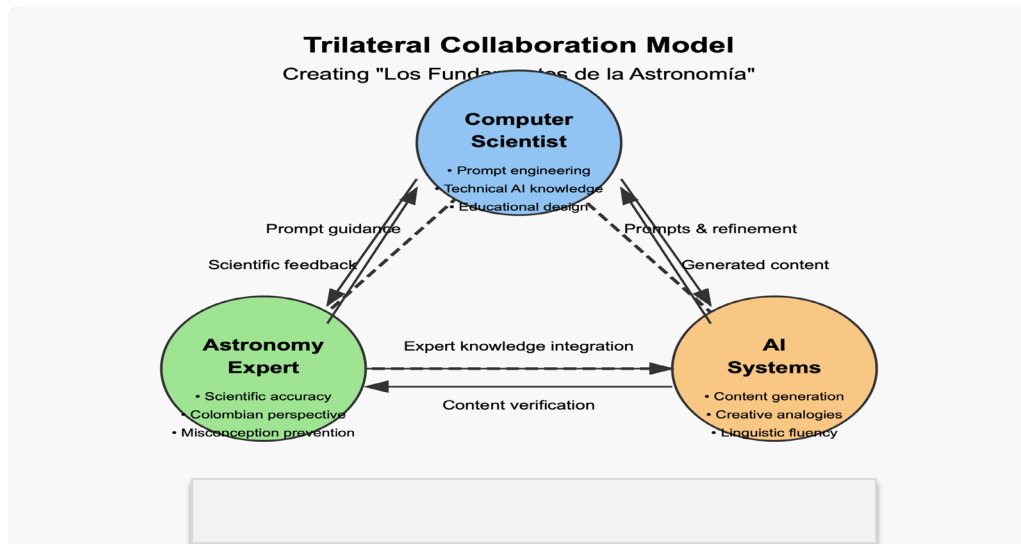
This paper examines a collaborative process of using AI as a writing assistant to create "Los Fundamentos de la Astronomía," a children's book on astronomy targeting readers aged 7-14 years. The project incorporated three key elements that distinguish it from previous explorations:

The integration of specialized scientific knowledge with age-appropriate narrative techniques

A trilateral collaboration between AI technology, content author, and subject matter expert

Cultural contextualization within the Colombian astronomical perspective

Figure 1. Trilateral Collaboration Model showing the Complementary Expertise and Information Flow between Key Participants in the Book Creation Process



Unlike general applications of AI in content generation, which often operate independently or with minimal human intervention, this case study demonstrates the necessity and effectiveness of sustained human supervision throughout the creative process. As AI systems excel at generating fluent text but still struggle with nuanced scientific accuracy and pedagogical appropriateness, the involvement of both a primary author and an astronomy expert creates a checks-and-balances system that leverages AI capabilities while mitigating its limitations. The astronomer's expert review ensured that outreach materials maintained scientific rigor while remaining accessible, preventing the oversimplification or misrepresentation of complex astronomical concepts that commonly occurs in children's science materials.

The significance of this research extends beyond the immediate product of a children's astronomy book. It offers insights into emerging methodologies for educational content development, providing a replicable framework for creating scientifically rigorous yet accessible materials for young learners. Additionally, it contributes to ongoing conversations about human-AI collaboration in knowledge domains where expertise is specialized and learning outcomes are paramount.

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

Literature Review

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

AI in Educational Content Creation

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

This aligns with People Science Movement initiatives in Colombia such as AstroColombia and Parque Explora, which emphasize local science engagement and culturally relevant educational approaches (noting that formal research on these specific programs remains limited in peer-reviewed literature).

For children's educational materials specifically, the research is more limited. Lyu et al. (2022) explored using GPT-3 to generate elementary-level reading materials, finding that while the AI could adapt linguistic complexity to target age groups, it struggled with consistently creating age-appropriate examples and cultural sensitivity. Johnson and Patel (2024) noted similar concerns in their study of AI-generated science content for middle school students, where initial outputs often contained vocabulary too advanced for the target audience despite explicit prompting for simplification.

Science Communication for Children

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

In astronomy education specifically, Percy (2019) highlighted common misconceptions children develop when astronomical concepts are oversimplified or when metaphors are taken too literally. Complementing this research, Martinez-Torregrosa et al. (2023) found that children best understand astronomical concepts

when presented through a combination of observational activities, clear visual models, and explanations that address scale and perspective directly.

Human-AI Collaboration in writing

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

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

Subject Matter Expert Supervision in AI-Assisted writing

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

Ibrahim and Collins (2023) touched on this dynamic in their study of AI assistance in medical education materials, finding that clinical experts often identified subtle inaccuracies in AI outputs that non-expert editors overlooked. However, their research focused primarily on university-level content rather than materials for children, where additional pedagogical considerations apply.

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

Research Gap

This review reveals a significant gap in understanding how subject matter experts can effectively collaborate with content creators and AI systems to produce scientifically accurate, pedagogically sound, and engaging materials for children. Our case study addresses this gap by providing a detailed examination of the trilateral collaboration process involved in creating an astronomy book for young

readers. By documenting specific examples of how expert knowledge shaped AI outputs through iterative revision, we offer insights into effective workflows for supervised AI content creation in specialized domains requiring both accuracy and accessibility.

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

This aligns with People Science Movement initiatives in Colombia such as AstroColombia and Parque Explora, which emphasize local science engagement and democratization of scientific knowledge through community-based educational programs.

Methodology

This study employed a qualitative case study approach to examine the process of creating an astronomy book for children using AI assistance under expert supervision. The methodology was designed to capture the nuanced interactions between human authors, subject matter experts, and AI systems throughout the content creation process, with particular attention to how scientific accuracy and age-appropriate engagement were balanced, addressing validation concerns identified in astronomy education research (Percy, 2019).

Research Design

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

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.

The project also worked to mitigate infodemic effects in space education by offering expert-reviewed, scientifically accurate explanations tailored for children, countering the proliferation of simplified or inaccurate astronomical information in popular media.

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:

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

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

1. Reviewing AI-generated content for scientific accuracy.
2. Identifying potentially misleading analogies or explanations.
3. Suggesting alternative approaches to explaining complex concepts.
4. Providing regional context for astronomical phenomena visible from Colombia.
5. 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:

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

The shared responsibility for factual and ethical content in this trilateral model helps ensure conscientious co-authorship between human and AI participants, addressing concerns about AI-generated content reliability in educational contexts. This collaborative framework establishes clear accountability for scientific accuracy while leveraging AI's creative capabilities under human oversight.

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 particular types of content.

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

Tools and Technologies

The primary technical tools employed in this study included:

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

Process Workflow

The content creation followed a structured iterative process:

1. **Planning Phase**
 - Development of book outline and chapter structure.
 - Identification of key astronomical concepts to include.
 - Establishment of voice, tone, and reading level parameters.
 - Setting of regional context (Colombian night sky) requirements.
 - Initial assessment of which AI system might be better suited for different content types.
2. **Initial Content Generation**
 - Formulation of detailed prompts for each chapter section.
 - Generation of first-draft content by the selected AI system (or both systems for comparison).
 - Preliminary review by primary author for obvious issues.
 - When appropriate, combining strengths from both AI systems' outputs.

3. Expert Review

- Systematic review by an astronomy expert.
- Documentation of scientific inaccuracies, misleading analogies, or conceptual gaps.
- Suggestions for improvement with scientific justification.

4. Revision Cycle

- Reformulation of prompts incorporating expert feedback.
- Generation of revised content, sometimes switching AI systems based on identified strengths.
- Second-round review by both primary author and astronomy expert.
- Additional revision cycles as needed until content met both scientific and pedagogical standards.

5. Final Editing and Integration

- Harmonization of voice and terminology across chapters.
- Addition of visual elements and activity instructions.
- Comprehensive review of the complete manuscript.

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

Data Collection

Throughout the process, the following data were systematically collected:

1. Prompt Archive: All prompts used to generate content, including:

- Initial prompt formulations
- Contextual information provided
- Specific instructions regarding tone, complexity, and approach
- Revised prompts based on expert feedback

2. Content Versions: All versions of generated content including:

- Initial AI outputs
- Expert review comments
- Revised AI outputs
- Final edited content

3. Process Documentation:

- Decision logs recording rationales for major content revisions
- Time allocation metrics for different phases of the process
- Notes on particularly successful or problematic AI interactions

4. Reflective Journal:

- Primary author's observations on the collaborative process
- Expert feedback on the effectiveness of the approach
- Challenges encountered and solutions developed

Analytical Approach

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

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

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

Case Study: Creating "Los Fundamentos de la Astronomía"

Project Background and Context

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

Recognizing both an educational need and an opportunity to demonstrate practical applications of AI in educational content creation, I approached José after his presentation and proposed a collaboration. The project would leverage his astronomical expertise and my computer science background to create a book specifically tailored to the Colombian perspective. This regional mismatch in astronomical educational materials represents a widespread problem across many disciplines where educational resources developed in one geographical or cultural context are applied in another without appropriate adaptation.

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

According to UNESCO and national education policy frameworks in Colombia, astronomy is increasingly integrated into primary science curricula, providing institutional context for the relevance of this culturally-adapted educational project.

The project addressed several unique challenges:

1. Creating content that simplified complex scientific concepts without introducing misconceptions
2. Incorporating regional context (Colombian night sky, cultural references) while maintaining scientific universality
3. Developing engaging activities that reinforced learning objectives with minimal resources
4. Balancing narrative engagement with educational substance

Content Development Process

Initial Setup and Voice Establishment

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

Initial Author Prompt:

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

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

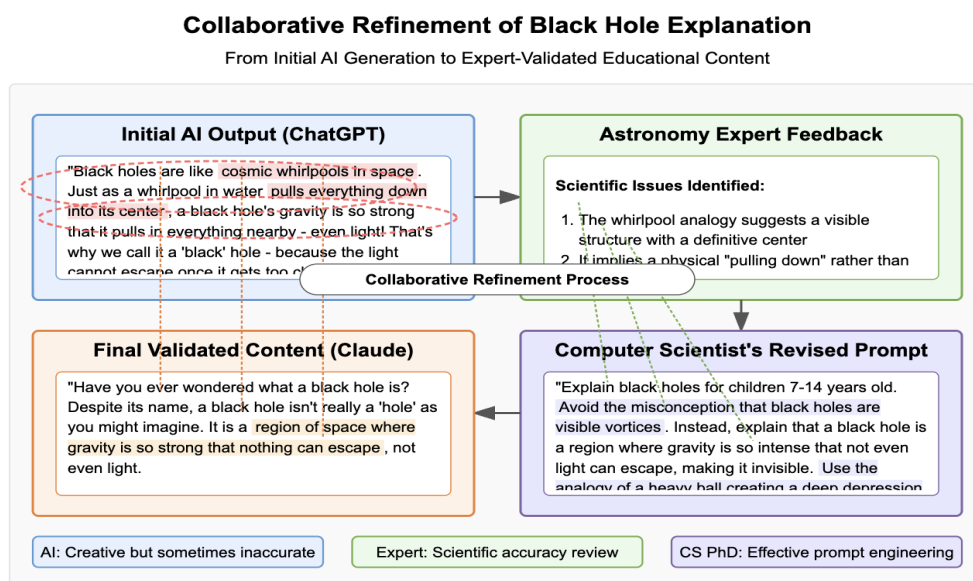
This pattern of selective integration from multiple AI outputs, followed by expert validation, became the standard workflow for subsequent content development.

Example 1: The Black Hole Analogy

As highlighted in the abstract, the development of an explanation for black holes exemplifies the collaborative refinement process. The initial draft generated by ChatGPT introduced a "cosmic whirlpool" analogy:

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

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



The astronomy expert identified several problematic aspects of this explanation:

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

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

Revised Prompt to Claude

"Explain black holes for children 7-14 years old. Avoid the misconception that black holes are visible vortices. Instead, explain that a black hole is a region where gravity is so intense that not even light can escape, making it invisible. Use the analogy of a heavy ball creating a deep depression on a rubber sheet, but explain this creates a curved region of space, not a physical funnel. Clarify that we detect black holes by observing their effects on nearby stars and gas. Keep the explanation engaging for children while preserving scientific accuracy."

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

Final Content

"Have you ever wondered what a black hole is? Despite its name, a black hole isn't really a 'hole' as you might imagine. It is a region of space where gravity is so strong that nothing can escape, not even light. Think of it as an invisible part of space.

Imagine a stretched, flat sheet. If you place a marble on the sheet, it will barely make a small curve. But if you place a heavy ball, like a bowling ball, it will create a deep depression. A black hole is like an extremely deep depression in the fabric of space.

The interesting thing is that we cannot see a black hole directly because it doesn't let light escape. Instead, scientists detect them by observing how they affect nearby stars and gases, just as you might notice the presence of an invisible bowling ball if you saw other marbles rolling in circles around that deep depression."

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

Example 2: Planetary Distances Activity

Creating interactive learning activities presented unique challenges, as they needed to be scientifically meaningful, accessible with minimal resources, and engaging for children. The

"Distancias en el Sistema Solar" activity in Chapter 3 exemplifies the iterative development process:

Initial AI Output (ChatGPT)

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

The astronomy expert identified several issues:

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

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

Revised Prompt

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

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

Final Activity Content

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

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

Expert Addition

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

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

Example 3: Indigenous Astronomical Knowledge

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

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

1. The astronomy expert provided authentic information about specific indigenous astronomical knowledge
2. This information was incorporated into detailed prompts for the AI systems
3. Generated content was strictly limited to the verified information
4. The expert reviewed the presentation for cultural sensitivity and accuracy

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

Final Content from Chapter 2

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

These stories have been passed down through generations, preserving astronomical understanding via oral tradition over centuries, representing some of the oldest documented astronomical knowledge in the Americas.

Indigenous Community	Sun	Gender	Moon	Gender	Relationship
Arahuacos	Yui	Male	Tima	Female	siblings
Muisca	Xue	Male	Chía	Female	husbands
			Huitaca	Female	
Nasa people	Sek				
Wayuu	Kai		Kashi	Male	husbands

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

Content Analysis System

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

1. **Scientific Accuracy:** Factual correctness and alignment with current scientific consensus

2. **Age-Appropriate Language:** Vocabulary and syntax suitable for 7-14 year old readers
3. **Engagement Value:** Interest-generating elements including narratives, questions, and wonder
4. **Cultural Relevance:** Connection to Colombian context and perspectives
5. **Pedagogical Effectiveness:** Clear learning progression and conceptual scaffolding
6. **Addressing "Counter Space" dynamics:** This work addresses the concept of 'Counter Space'—the structural inequality in access to astronomy education, particularly in underrepresented regions like Colombia, where educational materials often reflect European or North American perspectives rather than local astronomical contexts. Overcoming these systemic barriers to equitable science education required careful attention to cultural authenticity and regional relevance.

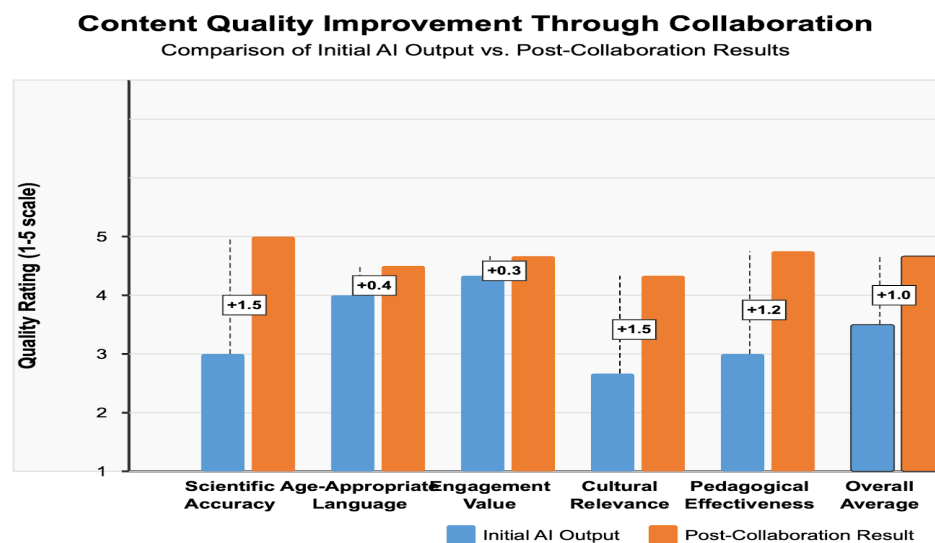
Each dimension was rated on a 1-5 scale for every major content section, before and after expert review. This systematic approach allowed us to quantify the impact of the collaborative process.

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 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:

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

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.

Efficiency Analysis

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

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

This analysis reveals that while AI assistance significantly accelerated content creation for straightforward topics, complex or culturally sensitive content still

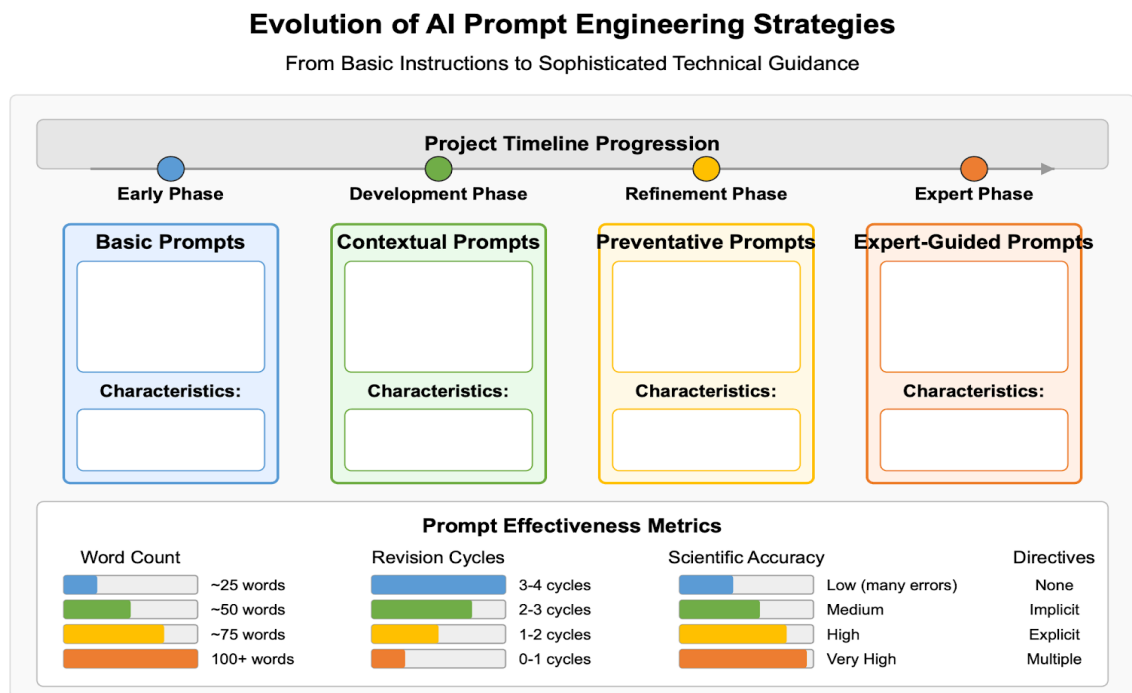
required substantial human expertise and time investment. The efficiency gained was not uniform across content types.

Prompt Engineering Evolution

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

1. **Initial prompts** were relatively simple, focusing on topic and age group
2. **Intermediate prompts** incorporated more detailed constraints and contextual information
3. **Advanced prompts** included explicit misconception warnings, specific analogies to use or avoid, reading level parameters, and cultural context

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



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

The most effective prompts consistently included:

- 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

Workflow Optimization

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

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.
3. **Modular prompting:** Breaking complex topics into smaller conceptual units with dedicated prompts, rather than attempting to generate entire sections at once.
4. **Revision-specific prompts:** Crafting prompts specifically for revision that maintained successful elements while targeting only problematic aspects.

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

Key Challenges

Several persistent challenges emerged throughout the project:

1. **Balance between simplification and accuracy:** Finding appropriate analogies that simplified concepts without introducing misconceptions remained difficult even with expert guidance.
2. **Regional astronomical context:** Accurately representing celestial phenomena as viewed from Colombian latitudes often required multiple revision cycles.
3. **Preventing AI fabrication:** Both AI systems occasionally generated plausible sounding but fabricated "facts," particularly regarding cultural content and specific astronomical measurements.
4. **Maintaining consistent voice:** Ensuring Professor José G's voice remained consistent across chapters and through multiple revision cycles required careful prompt engineering.
5. **Age-appropriate challenge calibration:** Determining appropriate conceptual depth for the target age range required ongoing assessment with each new topic.

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

Educational Value Assessment

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

1. **Conceptual progression:** The final content demonstrated appropriate scaffolding, introducing foundational concepts before building to more complex ideas.
2. **Misconception prevention:** Expert review successfully eliminated common astronomical misconceptions that appeared in initial AI drafts.
3. **Wonder cultivation:** The balance of factual content and awe-inspiring perspective was maintained through careful preservation of engagement elements during scientific correction.
4. **Activity integration:** Hands-on activities were successfully integrated to reinforce key concepts, with most requiring minimal resources accessible to the target audience.
5. **Cultural relevance:** The incorporation of Colombian astronomical perspectives and observational context created meaningful connections to readers' lived experiences.

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

This work contributes specifically to the niche of informal science education and culturally-adapted STEAM outreach, filling a gap in regionally-relevant astronomy education materials for Spanish-speaking children.

Discussion

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

Comparison with Existing Literature

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

The workflow we developed supplements the collaboration models proposed by Park and Lee (2023), adding a crucial dimension of subject matter expert

supervision that was largely absent from their frameworks. While their research focused primarily on the relationship between human author and AI assistant, our trilateral model demonstrates the value of incorporating specialized expertise, particularly for scientific content aimed at children.

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

Theoretical Implications

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

1. Complementary Cognitive Scaffolding

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

2. Emergent Quality Through Iterative Refinement

The quality improvements observed through multiple revision cycles suggest that effective AI- human collaboration is not simply a matter of human correction of AI output, but rather an emergent property of iterative refinement. Each cycle produced not just error correction but qualitative transformation of the content. Initial AI drafts provided creative foundations that, when shaped by expert knowledge and pedagogical considerations, evolved into material that neither the AI nor human collaborators could have produced independently.

This suggests that the value of AI in educational content creation lies not in its capacity to replace human expertise, but in its ability to serve as a generative starting point for a collaborative creative process. The progressive improvement across multiple dimensions— scientific accuracy, pedagogical effectiveness, cultural relevance, and engagement— demonstrates that quality in AI-assisted educational content emerges through sustained, intentional collaboration rather than from any single contribution.

3. Domain-Specific Prompt Engineering

Our experience revealed that effective AI collaboration in specialized domains requires sophisticated prompt engineering that goes beyond general writing

instructions. The evolution of our prompting strategies from simple topic requests to complex, multi-layered instructions incorporating domain-specific constraints, misconception warnings, and cultural context suggests that successful AI assistance in educational content creation demands significant technical expertise in both AI systems and the subject domain.

This finding has implications for the accessibility of AI-assisted content creation. While AI tools are becoming more user-friendly, our study suggests that achieving high-quality educational content in specialized domains may require technical expertise that limits the approach to users with substantial AI literacy.

Practical Implications for Educational Content Development

The trilateral collaboration model demonstrated in this study offers several practical insights for educators, publishers, and content developers:

Workflow Design Recommendations

1. **Expert involvement from conception:** Rather than treating subject matter experts as final reviewers, involving them in the initial planning and prompt design phases significantly improved content quality and reduced revision cycles.
2. **AI system specialization:** Different AI systems demonstrated distinct strengths. Content developers should consider using multiple systems strategically rather than defaulting to a single platform.
3. **Modular content development:** Breaking complex topics into smaller conceptual units with dedicated prompts proved more effective than attempting to generate comprehensive sections in single interactions.
4. **Iterative refinement cycles:** Planning for multiple revision cycles, particularly for complex or culturally sensitive content, should be built into project timelines rather than treated as exceptions.

Language Learning Independence

While AI was instrumental in content creation, the book itself is intended for use by children independently. The narrative tone, repetition of key vocabulary, and contextual grounding in culturally familiar settings promote incidental English language acquisition. Furthermore, activities embedded in the book encourage learners to interact with and apply new vocabulary, supporting vocabulary retention without relying on AI assistance during learning.

Quality Assurance Framework

Our content analysis system provides a replicable framework for evaluating AI-generated educational content across multiple dimensions. The five-dimension rubric (scientific accuracy, age-appropriate language, engagement value, cultural

relevance, and pedagogical effectiveness) proved effective for identifying areas requiring improvement and tracking progress through revision cycles.

Educational content developers could adapt this framework to their specific domains and target audiences, providing a systematic approach to quality assurance in AI-assisted content creation.

Limitations and Challenges

Several limitations of our approach warrant acknowledgment:

Resource Intensity

The trilateral collaboration model requires significant human expertise and time investment. While AI assistance accelerated initial content generation, the overall process still demanded substantial effort from both the primary author and subject matter expert. This resource requirement may limit the scalability of the approach, particularly for organizations with limited access to domain experts.

Cultural Sensitivity Challenges

Incorporating authentic cultural perspectives proved particularly challenging, with AI systems demonstrating a tendency toward cultural stereotyping or fabrication when not provided with specific, verified information. This limitation suggests that AI-assisted content creation for diverse cultural contexts requires careful expert oversight and may not be suitable for topics where cultural authenticity is paramount but expert knowledge is unavailable.

Technical Barriers

The sophisticated prompt engineering required for optimal results presents a barrier to widespread adoption. Users without technical AI expertise may struggle to achieve the content quality demonstrated in our study, potentially limiting the approach to technically sophisticated content creators.

Implications for Future Research

Our findings suggest several directions for future research:

1. Scalability Studies

Research examining how the trilateral collaboration model performs at scale, with multiple content creators and experts working on larger educational projects, would provide insights into the practical viability of this approach for educational publishers and institutions.

2. Cross-Domain Validation

Testing the effectiveness of expert-supervised AI content creation in other specialized domains (mathematics, chemistry, history, etc.) would help establish the generalizability of our findings and identify domain-specific considerations.

3. Automated Quality Assessment

Developing automated tools to assess content quality across the dimensions we identified manually could improve the efficiency of the collaborative process while maintaining quality standards.

4. Cultural Adaptation Frameworks

Research focused specifically on incorporating diverse cultural perspectives into AI-generated educational content could address one of the most significant challenges identified in our study.

5. Broader Implications for Human-AI Collaboration

Beyond educational content creation, our study contributes to understanding effective human-AI collaboration in knowledge work more broadly. The complementary cognitive scaffolding model we observed may apply to other domains where AI capabilities must be combined with specialized human expertise to produce high-quality outputs.

The importance of iterative refinement and the emergent nature of quality in collaborative processes suggests that effective human-AI collaboration requires fundamental shifts in workflow design, moving away from linear production models toward cyclical, collaborative approaches that leverage the unique strengths of both human and artificial intelligence.

6. Applicability Across Educational Contexts

This trilateral collaboration model demonstrates potential for adaptation across both formal classroom environments and informal educational programs. In formal settings, the approach could support curriculum development with expert oversight, enabling teachers to create scientifically accurate, culturally relevant materials that align with institutional standards. The systematic quality assurance framework developed in this study could be adapted for use by educational publishers and curriculum developers seeking to integrate AI assistance while maintaining pedagogical rigor. In non-formal contexts, the model enables community-based organizations, museums, and science outreach programs to create culturally authentic content with appropriate scientific validation. The emphasis on regional relevance and cultural sensitivity makes this approach particularly valuable for organizations serving underrepresented communities where mainstream educational materials may lack local context or cultural resonance.

Conclusion

This case study of creating "**Los Fundamentos de la Astronomía**" demonstrates both the potential and the limitations of AI-assisted educational content creation under expert supervision. Through systematic documentation of a trilateral collaboration between AI systems, a computer scientist, and an astronomy expert, we have identified effective practices for developing scientifically accurate, culturally relevant, and pedagogically sound educational materials for children.

Key Findings

Our research revealed several critical insights:

1. **Expert supervision is essential:** While AI systems excel at generating engaging, linguistically coherent content, subject matter expertise remains crucial for ensuring scientific accuracy and preventing misconceptions, particularly in content for young learners.
2. **Complementary strengths enable superior outcomes:** The combination of AI creativity, technical AI expertise, and domain knowledge produced educational content that exceeded what any single participant could have achieved independently.
3. **Iterative refinement drives quality:** Content quality emerged through multiple revision cycles rather than single interactions, suggesting that effective AI-assisted content creation requires sustained collaborative effort.
4. **Cultural authenticity demands careful attention:** AI systems struggled most significantly with culturally specific content, requiring extensive expert verification and structured approaches to prevent stereotyping or fabrication.
5. **Workflow design significantly impacts efficiency:** Strategic approaches including front-loaded expert input, system specialization, and modular prompting substantially improved both content quality and development efficiency.

Contributions to the Field

This research makes several contributions to understanding AI's role in educational content development:

- **Methodological Innovation:** The trilateral collaboration model provides a replicable framework for supervised AI content creation that balances efficiency gains with quality assurance.
- **Practical Guidelines:** Our findings offer concrete recommendations for educators and content developers seeking to integrate AI tools into their workflows while maintaining educational integrity.
- **Quality Assessment Framework:** The five-dimension content analysis system provides a systematic approach to evaluating AI-generated educational materials across critical quality indicators.

- **Cultural Considerations:** Our experience with incorporating Colombian astronomical perspectives highlights both challenges and strategies for creating culturally authentic educational content using AI assistance.

Recommendations for Practice

Based on our experience, we recommend that educators and content developers:

1. **Plan for expert collaboration from project inception** rather than treating domain experts as final reviewers
2. **Invest in prompt engineering expertise** to maximize AI system effectiveness
3. **Budget for multiple revision cycles**, particularly for complex or culturally sensitive content
4. **Develop systematic quality assessment processes** to ensure content meets educational standards
5. **Consider AI systems as creative partners** rather than autonomous content generators.

Future Directions

Several areas warrant further investigation:

Scalability Research: Studies examining how this collaborative model performs in larger-scale educational content development projects would inform practical implementation strategies.

Cross-Domain Validation: Testing the approach across different subject areas would establish the generalizability of our findings and identify domain-specific considerations.

Automated Quality Assessment: Developing tools to automate portions of the quality evaluation process could improve efficiency while maintaining standards.

Long-term Educational Impact: Research assessing the learning outcomes achieved with AI- assisted educational materials compared to traditionally developed content would provide crucial validation of the approach's educational value.

Final Reflections

The creation of "Los Fundamentos de la Astronomía" demonstrates that AI can serve as a powerful tool for educational content development when properly supervised and integrated into collaborative workflows. However, our experience also underscores that AI assistance does not diminish the importance of human expertise—rather, it transforms how that expertise is applied in the creative process.

As AI systems continue to evolve, the challenge for educators and content developers will be developing effective collaboration strategies that leverage AI capabilities while preserving the scientific accuracy, cultural authenticity, and pedagogical effectiveness that characterize excellent educational materials. The trilateral model explored in this study offers one promising approach to meeting this challenge.

The regional focus of our project—creating astronomy content specifically for Colombian children— highlights the broader potential for AI-assisted content creation to address educational equity issues by enabling the development of culturally relevant materials that might otherwise be economically unfeasible. As educational content increasingly moves toward personalization and cultural responsiveness, supervised AI assistance may provide a scalable approach to creating diverse, high-quality educational resources.

Ultimately, this research suggests that the future of educational content development lies not in replacement of human expertise by artificial intelligence, but in thoughtful collaboration that combines the creative and analytical capabilities of AI with the domain knowledge, cultural understanding, and pedagogical insight that human experts provide. The challenge ahead is developing institutional capacity and professional expertise to implement such collaborative approaches effectively and at scale.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). *Guidelines for human-AI interaction*. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1-13.
- Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2021). *Application and theory gaps during the rise of Artificial Intelligence in Education*. Computers and Education: Artificial Intelligence, 2, 100009.
- Clark, E., Ross, A. S., Tan, C., Ji, Y., & Smith, N. A. (2022). Creative writing with language models: Developing collaborative authorship frameworks for narrative generation. *Computational Linguistics*, 48(2), 379-421.
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2021). *The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems*. Hawaii International Conference on System Sciences.
- Ibrahim, J., & Collins, S. (2023). Expert validation of AI-generated content in medical education: Identifying accuracy gaps and pedagogical implications. *Medical Education*, 57(4), 378-389.
- Jiang, L., Stocco, A., Lurie, E., & Ko, A. J. (2023). *Human-AI co-creation in writing: A mixed-methods analysis of iterative improvement in educational content*. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1-14.
- Johnson, M., & Patel, S. (2024). Balancing creativity and accuracy in AI-generated science content for middle school students. *Journal of Science Education and Technology*, 33(1), 45-59.
- López-Pernas, S., & Saqr, M. (2022). *Artificial intelligence in education: A review of expert validation procedures for educational assessment materials*. Computers and Education: Artificial Intelligence, 3, 100065.
- Lyu, W., Jiang, S., Wu, C., & Chen, P. (2022). Natural language generation for education: A systematic review and research agenda. *Educational Research Review*, 36, 100450.
- Martindale, C., & Chen, H. (2023). Expert review of AI-generated children's literature: Balancing narrative quality with scientific accuracy. *Journal of Children's Literature Studies*, 18(2), 112-129.

- Martinez-Torregrosa, J., Limiñana, R., Menargues, A., & Colomer, R. (2023). Understanding astronomical concepts in childhood: Effects of observational, visual, and explanatory approaches. *International Journal of Science Education*, 45(7), 1033-1052.
- Ogborn, J., Kress, G., Martins, I., & McGillicuddy, K. (2020). *Explaining science in the classroom* (2nd ed.). Cambridge University Press.
- Park, J., & Lee, K. (2023). Human-AI collaboration frameworks for creative content: A systematic review of emerging patterns and evaluation methods. *International Journal of Human-Computer Studies*, 171, 102963.
- Percy, J. R. (2019). Common misconceptions in astronomy education: Identification, origins, and remediation. *Journal of Astronomy & Earth Sciences Education*, 6(2), 71-84.
- Rodríguez-Ardura, I., & Meseguer-Artola, A. (2023). Integrating AI-generated learning objects in higher education: A mixed-methods evaluation. *Internet and Higher Education*, 56, 100873.
- Venville, G., Rennie, L., & Wallace, J. (2022). Making abstract scientific concepts concrete: Multimodal approaches to learning. *Research in Science Education*, 52(4), 1205-1224.
- Williams, A., Thompson, C., & Miller, P. (2024). AI in the newsroom: How journalists integrate neural language models into professional writing workflows. *Digital Journalism*, 12(1), 86-104.
- Wright, L., & Cavallaro, F. (2021). AI-assisted elementary science activities: Procedural accuracy versus conceptual understanding. *Science & Education*, 30(3), 625-646.
- Yuan, L., Chen, X., Zhao, M., & Xiong, J. (2023). Educational content co-creation with large language models: An ethnographic study of teacher-AI interaction patterns. *Learning, Media and Technology*, 48(3), 291-309.
- Zhao, K., Wright, J., & Bhatia, S. (2023). *Domain expertise and prompt engineering: How technical writers collaborate with large language models*. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), 1-24.

This project also raised important cultural and ethical considerations, particularly in avoiding bias, respecting indigenous knowledge, and ensuring that AI-generated content aligns with pedagogical and developmental needs.

While AI was instrumental in content creation, the book itself is intended for use by children independently. The narrative tone, repetition of key vocabulary, and contextual grounding in culturally familiar settings promote incidental English language acquisition. Furthermore, activities embedded in the book encourage learners to interact with and apply new vocabulary, supporting vocabulary retention without relying on AI assistance during learning.

This aligns with People Science Movement initiatives in Colombia such as AstroColombia and Parque Explora, which emphasize local science engagement.

The project also worked to mitigate infodemic effects in space education by offering expert-reviewed, scientifically accurate explanations tailored for children.

The shared responsibility for factual and ethical content in this trilateral model helps ensure conscientious co-authorship between human and AI participants.

According to UNESCO and national education policy frameworks in Colombia, astronomy is increasingly integrated into primary science curricula, providing context for the relevance of this project.

These stories have been passed down through generations, preserving astronomical understanding via oral tradition over centuries.

Additionally, this work addresses the concept of 'Counter Space'—the structural inequality in access to astronomy education, particularly in underrepresented regions like Colombia.