

1. Project Summary

Problem & Motivation - In complex academic writing, users often suffer from "cognitive fixation" - early commitment to a single idea or framing. Dominant chat-based LLMs frequently reinforce this issue by optimizing for the most probable continuation, encouraging passive consumption and reducing users to mere editors of generic, AI-generated text rather than active participants in the reasoning process.

Solution: **Lantern** is a "Thinking Partner" designed to shift interaction from passive generation to active exploration. The system employs a Tree-of-Thoughts (ToT) framework, externalizing reasoning into a persistent, navigable visual map. This allows users to track their exploration history and compare parallel arguments while ensuring synthesis and authorship remain entirely human-driven.

Unique Approach: Bridging "Wants" vs. "Needs" - Lantern prioritizes cognitive needs over convenience through specific design constraints:

- **Productive Friction:** The system requires explicit selection between alternatives, shifting users from autopilot behavior toward analytical System-2 thinking.
- **Challenge over Validation:** A dedicated "Devil's Advocate" agent surfaces logical gaps and counter-arguments to counter AI sycophancy.
- **Conceptual Anchors:** AI outputs serve as conceptual guides in the sidebar and are never directly injected into the document, preserving the user as the primary architect of the text.
- **Controlled Refinement:** To maintain a seamless workflow, the **Refine** agent polishes academic style and clarity upon request without altering core ideas. This integrated feature eliminates the need for external tools like Grammarly while keeping the user's original meaning intact.
- **Embedded Expertise:** By encoding standards derived from **90 academic writing sources** directly into agent roles, the system provides expert-level guidance and professional prompt engineering without increasing user cognitive load.

Lantern's true value lies in stimulating diverse ideas and logical precision rather than word count, positioning the human as the master architect of every thought.

In case you encounter a problem accessing the website, you can view and run the code from our [Git repository](#)

Interface - The interface is designed to transform the academic writing process from linear text generation into a dynamic exploration of ideas. By integrating directly as a sidebar within a document editor, it serves as a conceptual guide rather than a content replacement tool.

Core Components & Supported Tasks

- **Integrated text editor:** A familiar environment where users perform the primary tasks of synthesis and authorship.
- **Reasoning map (ToT):** A visual map that displays the reasoning process, allowing users to navigate parallel lines of thought.
- **Academic Justification Labels:** Every AI-generated output is accompanied by a clear explanation that identifies the specific academic principle (e.g., logical coherence, evidence-based claims) used to generate that path.
- **Refine agent:** A specialized agent that improves text clarity and academic tone, enabling users to polish their writing within the system without needing external tools like Grammarly. This tool is specifically designed to refine the presentation and style of the draft while ensuring the user's original ideas, core arguments, and underlying meaning remain entirely unchanged.
- **Critique Agent** - This agent is designed to simulate a critical research peer or a strict academic reviewer. Its role is identifying weaknesses in argumentation and checking for logical consistency. The agent scans the selected text and looks for: logical fallacies, counter-arguments, lack of evidence and confirmation bias. Instead of altering the text, the agent presents Annotations or that ask difficult questions, allowing the user to choose whether to correct the argument or sharpen it.
- **Argument StrengthTracker:** Designed to overcome the natural human resistance to criticism, this feature transforms feedback into a measurable asset. By "Acknowledging" (👉) valid critiques, users build a visual strength score in their dashboard. This approach motivates users to embrace difficult feedback, turning the resolution of logical gaps into tangible proof that their thesis is becoming bulletproof.
- **Idea Expander** - This agent is designed to assist during moments of "writer's block" or when the discussion feels too shallow. It operates on the principle of Divergent Thinking - generating a variety of ideas from a single starting point. Its primary role is opening new lines of thought and deepening the academic discussion. How it works: The agent takes an idea, sentence, or paragraph and suggests:
 - **Alternative Perspectives:** How does the issue look from the viewpoint of another discipline (e.g., a sociological aspect of a technological problem)?
 - **Deepening:** Proposing additional research questions that arise from the current text but remain unanswered.
 - **Implications:** What are the future or theoretical implications of the claim being made?
 - **Examples and Analogies:** Suggesting similar cases that can illustrate the abstract idea.

The Output is a list of possible Thought Paths. The user can select one, and the system will create a new branch in the Reasoning Map focused on this new direction, without deleting the original idea.

- **Paragraph Selection** - The system automatically detects and segments paragraphs within the editor. This allows users to submit specific text blocks to the agents, rather than the entire document, ensuring that AI outputs (whether refining, critiquing, or expanding) remain focused on the immediate context the user wishes to explore.
- **Current Focus Area (Pinned Context)**: A space at the top of the sidebar where users anchor specific ideas or critiques to help focus their writing process.

HAI Guidelines

The system's architecture is grounded in three core HAI guidelines:

1. **Balance Automation & Augmentation**: The system favors a "Thinking Partner" approach over full automation. The AI augments human reasoning by expanding the solution space, while the user remains the primary decision-maker.
2. **N-best Alternatives**: To mitigate bias and build trust, the interface presents multiple distinct branches. To enhance explainability, each alternative is explicitly linked to one of the academic standards encoded in the system, helping users understand the logic behind every suggestion.
3. **Offer Control**: Users steer the AI through explicit control mechanisms, ensuring the output aligns with their immediate intent

System Enhancements - Significant changes were made to the prototype based on feedback received during the project presentation in class:

- **Tree-of-Thoughts Visualization**: We introduced a visual tree to replace linear history, allowing users to see their entire exploration path.
- **Node-Based Selection**: We added a "**Select**" button that allows users to formally add reasoning paths from the Idea Expander as new nodes in the tree. These selected nodes are then fed back to the LLM as context for all subsequent outputs to ensure continuity. Conversely, the system remembers paths the user has chosen to **Prune**, ensuring they are no longer suggested in future interactions.
- **Enhanced Focus Control**: A "**Pin**" button was implemented for both ideas and critiques, allowing users to anchor specific LLM responses to the top of the sidebar. This ensures relevant insights remain visible to guide the user's drafting.
- **Project Management**: Functionality to import and export files was added to support multi-session work.
- **Standardized Text Editing**: We integrated basic text editing options (similar to popular tools like Word or Google Docs), providing a familiar and efficient environment for professional writing.

4.Theoretical Basis

Tree-of-Thoughts and Maximization Bias - Large Language Models are autoregressive next-token predictors, a property that often leads to maximization bias-early convergence on high-probability, generic continuations that reduce diversity of thought ([Holtzman et al., 2019](#); [Brown et al., 2020](#)). Yao et al. (2023) introduce the Tree-of-Thoughts (ToT) framework as a way to counter this limitation by explicitly maintaining multiple parallel reasoning paths instead of a single linear chain-of-thought. Lantern directly implements this theory at the interaction level: rather than presenting one response, the system externalizes several competing reasoning paths as separate, navigable artifacts. This design prevents premature

convergence and encourages users to actively compare alternatives instead of accepting the model's most likely continuation.

Dual Process Theory and Automation Bias - According to Dual Process Theory, human cognition alternates between fast, intuitive System 1 thinking and slow, analytical System 2 thinking ([Kahneman, 2011](#)). In Human-AI interaction, systems that provide fluent and authoritative outputs often trigger automation bias, leading users to default to System 1 acceptance of AI suggestions ([Chiang et al., 2024](#)). Lantern is explicitly designed to shift users toward System 2 engagement. This is achieved by requiring users to select, reject, or expand reasoning paths before any progression occurs, and through a Devil's Advocate agent that systematically challenges assumptions. In this way, the system does not automate reasoning but augments it, aligning with research that emphasizes augmentation over automation in cognitively demanding tasks.

Sycophancy, Agency, and Creativity Support Tools - Recent research highlights sycophancy in conversational AI systems—the tendency to align with user beliefs even when they are incorrect or weakly supported ([Sponheim, 2024; NNG, 2024](#)). Creativity Support Tools (CST) literature argues that effective systems should preserve user agency by allowing exploration, comparison, and intentional choice rather than passive acceptance ([Shneiderman et al., 2006](#)). Lantern operationalizes these principles by separating generation from decision-making: AI agents propose alternatives, while users retain full control through pinning and pruning mechanisms. This ensures that the final outcome reflects the user's intent rather than the model's tendency to please, establishing a shared mental model between human and AI.

3. User evaluation (human users):

To assess the effectiveness of Lantern in real-world academic writing, we conducted a qualitative user study with five participants (undergraduate students). Each participant was assigned a specific writing scenario and asked to complete it using the system's AI features.

Participant Scenarios

User	Scenario & Task	Feedback	Lessons Learned
U1	Argument Development: Expanding a short academic claim into a multi-faceted argument.	The user noted that when working on a full document, it was unclear which specific paragraph each AI suggestion referred to.	AI suggestions must be spatially anchored to specific text segments to reduce cognitive load and support informed decision-making.
U2	Relevance Assessment: Critiquing and strengthening the logic of an academic thesis.	While suggestions were linguistically high-quality, some were loosely connected to the specific	Linguistic quality is insufficient; prompt constraints must prioritize argumentative alignment and intellectual coherence over stylistic flair.

		context, leading to a risk of "AI hallucinations."	
U3	Refinement & Polishing: Performing a stylistic and structural revision of a full draft.	When refining the entire document, all improvements appeared as a single block, making it difficult to review changes individually.	Refinement feedback must be presented in a structured, navigable form (e.g., scrollable lists) explicitly linked to the relevant paragraphs.
U4	Reasoning Tree Navigation: Using the tree structure to explore and revisit different narrative paths.	The user struggled to recall the specific context or "reasoning" behind earlier nodes when returning to them after exploring other branches.	For tree-based reasoning to be effective, nodes must preserve provenance metadata, displaying the originating context and referenced text for each node.
U5	Targeted Feedback: Directing the AI to focus its analysis on a specific paragraph within a long text.	Manually identifying or counting paragraphs to guide the AI interrupted the user's "writing flow" and was prone to error.	Implementing automatic paragraph numbering and a refined "Focus Mode" is essential for seamless synchronization between user intent and AI output.

The evaluation highlighted that while the AI's generative capabilities are strong, the Human-AI collaboration relies on transparency and navigation. Based on these insights, we prioritized UI updates including paragraph-level indexing and context-tags for the Reasoning Tree to ensure that the system supports, rather than overwhelms, the user's intellectual process.

4. LLM-as-User Evaluation

We experimented with two complementary strategies for LLM-based evaluation:

1. LLM as a Design Critic (Theory-Aware Reviewer):

In this strategy, we used an LLM as a theory-aware design critic to evaluate Lantern's design approach at a conceptual level. The model was provided with a concise description of Lantern's interaction model, including Tree-of-Thought exploration, explicit reasoning modes, pinning, and the Devil's Advocate mechanism and asked to critically assess the design through established HCI/HAI lenses. Rather than evaluating outputs or usability details, the LLM was instructed to focus on design-level considerations, such as user agency, automation bias, cognitive load, and the balance between

augmentation and over-automation. The goal was to examine whether the proposed interaction logic coherently supports Lantern's stated design philosophy.

This strategy was used to surface implicit design assumptions, unresolved trade-offs, and potential tensions between theoretical intent and interaction structure. Importantly, the LLM served as a structured analytical lens for applying HAI principles consistently, rather than as an authority on design quality.

2. LLM-as-User Evaluation: A Design-Level Sanity Check:

To evaluate Lantern's design approach in a realistic way, we used an LLM to simulate a short interactive session with the system, rather than evaluating isolated outputs. The goal was to observe whether Lantern's interaction model consistently supports domain-relevant, non-hallucinatory, and non-sycophantic reasoning over the course of a dialogue.

The LLM was instructed to act as a domain-aware evaluator engaging in a simulated conversation with Lantern around a concrete academic topic. Instead of judging correctness, it was asked to reflect on the quality of the system's behavior: whether suggested directions meaningfully relate to the domain, whether they challenge assumptions rather than merely agreeing with them, and whether any responses appear fluent but epistemically shallow.

This conversational setup allowed us to assess Lantern's design tendencies such as how its expands, critique, and guidance mechanisms function together across turns, rather than relying on single, decontextualized examples.

Critique of LLM-Based Evaluation and Comparison with Human Feedback:

The LLM-based evaluation was useful primarily at the level of design principles and conceptual alignment, but was limited in identifying concrete interaction issues that emerged during testing with human users. Overall, the LLM's insights were not incorrect or unrealistic; however, they tended to be abstract and focused on potential usage patterns rather than immediate, practical difficulties.

When the LLM was used as a theory-aware design critic and as an evaluator across multiple interactions, it consistently identified that the system maintains a high level of user control over actions and decisions throughout the process. For example, it noted that action selection (Expand / Critique / Refine) is always explicit, that there are no default or ranked suggestions, and that the Refine mode requires separate approval for each proposed change. These observations support the claim that the design reduces automation bias and discourages passive acceptance of AI output.

In addition, the LLM identified stable behavioral patterns of the system from which it inferred potential long-term risks. Across repeated simulations and multiple domains, it observed that the Expand action alone tends to generate relevant but "safe" directions, and that only the combination with Critique introduces meaningful epistemic tension and challenges underlying assumptions. This suggests that, under repeated use, users who rely primarily on expansion without critique may remain within familiar conceptual framings despite apparent diversity.

In contrast, human users primarily reported concrete and immediate usability issues. They struggled to understand which paragraph a given piece of feedback referred to, experienced confusion when navigating back to earlier nodes in the reasoning tree without sufficient contextual cues, and found it difficult to direct the system's focus to a specific paragraph in longer documents. These issues recurred across multiple tasks and point to shortcomings in interaction clarity and context preservation, areas that the LLM largely did not surface.

Reflection on LLM use

During this project, our experience with LLMs varied significantly depending on the task at hand. Initially, we utilized Gemini to brainstorm various project concepts. However, this approach proved ineffective; the suggestions were repetitive and lacked genuine creativity, often cycling through generic ideas rather than offering novel solutions.

In contrast, our use of Perplexity for the research phase was highly successful. We used it to scan and synthesize approximately 90 sources regarding proper academic writing standards. This resulted in a deep analysis and a well-structured specification document for the LLM we eventually developed. The summarization of these sources was comprehensive, accurate, and provided a solid foundation for our work.

Throughout the execution phase, we used Gemini to refine the English grammar in our written submissions and to assist with coding. While helpful for small tasks, the coding assistance became cumbersome as the codebase grew in length and complexity. Gemini tended to omit sections of code or arbitrarily alter logic without being prompted. This required us to meticulously review every output to understand the changes, and we often found that manually modifying the code was faster and safer than relying on the LLM to handle complex files.

Our key learning is that while LLMs struggle with open-ended creative ideation, they excel at implementing specific, well-defined methodologies. For example, when we asked Gemini to apply Marty Cagan's product design methodology, it performed exceptionally well. We concluded that LLMs are powerful tools when you know exactly what you want and provide precise instructions; however, when left with too much room for interpretation, the quality of the output significantly declines.

Summary

Developing Lantern was a transformative experience that bridged the gap between theoretical Human-AI Interaction concepts and practical software engineering. On a technical level, We gained hands-on experience in orchestrating multi-agent systems using LLM APIs. We learned how to architect distinct personas, such as the "Devil's Advocate" and "Idea Expander", and coordinate their interactions to serve a unified user goal. This required deep engagement with advanced prompt engineering, specifically how to embed domain expertise (derived from academic literature) into system instructions to ensure high-quality, context-aware outputs.

Theoretically, we learned how to operationalize cognitive frameworks. We moved beyond simply understanding the "Tree-of-Thoughts" or "Dual Process Theory" to actually designing UI mechanisms, like the visual reasoning map and mandatory node selection, that enforce these principles.

Most significantly, this project instilled us in a strong sense of technological self-efficacy. Initially, the prospect of building a complex, AI-driven application from scratch felt daunting. However, seeing Lantern evolve from a concept to a working prototype has given us the confidence to architect and deploy our own independent projects. We have transitioned from consumers of AI tools to creators capable of shaping how humans interact with them.

Future Work & Extensions

With more time, we would significantly expand Lantern's capabilities to better support authentic academic workflows. A primary focus would be integrating Retrieval-Augmented Generation (RAG) connected to live databases like Google Scholar. This would allow the *Idea Expander* to perform Citation Injection, grounding its suggestions in actual literature to mitigate model hallucinations. To support the non-linear nature of research, we plan to implement a "Graveyard" feature for archiving pruned branches, enabling users to resurrect discarded ideas if their inquiry shifts. Furthermore, we aim to develop internal Fixation Metrics to track user pivoting behavior; if the system detects prolonged engagement with a single branch, it would proactively trigger the *Devil's Advocate* to challenge the user's focus. Finally, to ensure real-world utility, we would refactor Lantern from a standalone editor into a plugin for Microsoft Word or Google Docs, meeting researchers in their native environments while retaining the distinct reasoning map interface.

Appendix:

1. Prompts and Raw LLM Responses

Strategy 1: LLM as a Design Critic (Theory-Aware Reviewer)

Prompt:

"You are evaluating an AI system called Lantern.

Lantern is a reasoning-support tool rather than a conversational chatbot.

Users do not interact through free-form dialogue.

Instead, they select existing text and explicitly trigger reasoning actions.

Lantern provides three main actions:

- Expand: generate multiple alternative reasoning directions
- Critique: challenge assumptions, gaps, and framings (Devil's Advocate)
- Refine: improve clarity and structure without adding new ideas

The system is designed to preserve user agency, reduce automation bias, and support deliberate, reflective reasoning.

During this evaluation, you will be given:

1. Raw textual outputs produced by Lantern
2. Interface screenshots illustrating how actions are exposed to users

Your task is to evaluate whether Lantern's DESIGN APPROACH

supports interaction-level principles such as:

- user agency
- cognitive load management
- progressive disclosure
- resistance to automation bias

Do NOT critique visual aesthetics or UI polish.

Focus on interaction logic and behavioral implications."

Raw LLM Response:

"Iteration 1 — Visibility & User Agency

What you do in Lantern:

1. Select a short text (1–2 paragraphs).

2. Make sure the Expand / Critique / Refine actions are visible.

3. Press Expand on the selected text.

What you send me:

1. One screenshot showing:

- Expand / Critique / Refine actions
- How paragraph selection works

2. The raw Expand output, exactly as produced"

Prompt:

1.Expand:

The screenshot shows the Lantern AI interface. On the left, the 'Editor' section displays a document with the following text:
It is crucial for education to adjust to and adopt technological innovations in our constantly changing society. Artificial intelligence (AI) has become a potent technology with great potential for changing how we teach and learn. AI has the power to transform education, creating new opportunities and improving students' educational experiences.
AI in education delivers individualised learning opportunities that meet the specific needs of each student. Through intelligent algorithms and data analysis, AI can adapt content, pace, and instructional approaches to match students' unique learning styles and abilities. This personalised approach fosters engagement and boosts student achievement.

At the top of the editor, there are three main buttons: 'Expand', 'Critique', and 'Refine'. Below these are sections for 'AI Context & Structure' and 'Focus Range' (set to 'Specific Paragraph'). A sidebar on the right shows pinned context and suggested paths. The top right corner features a lantern icon and links for 'Import' and 'Export'.

Suggested Paths
Lantern generated alternative reasoning paths. Select one to continue.

[P2] A. Cognitive Load Theory Perspective
Focus: Paragraph 2

Explore how AI-driven personalization affects cognitive load. Does tailoring content reduce extraneous cognitive load, freeing up resources for germane cognitive load and deeper learning? Or does constant adaptation create a shifting landscape that increases cognitive load and anxiety? Cite relevant cognitive load theory research to support your analysis.

Standard [P1] B. Historical Precedent & Hype Cycle Analysis
Focus: Paragraph 1

"[P2] A. Cognitive Load Theory Perspective i Focus: Paragraph 2

Explore how AI-driven personalization affects cognitive load. Does tailoring content reduce extraneous cognitive load, freeing up resources for germane cognitive load and deeper learning? Or does constant adaptation create a shifting landscape that increases cognitive load and anxiety? Cite relevant cognitive load theory research to support your analysis.

standard

[P1] B. Historical Precedent & Hype Cycle Analysis i Focus: Paragraph 1

Frame AI within a history of technological panaceas in education (e.g., radio, film, computers). Map AI onto the Gartner Hype Cycle. Is AI currently overhyped in education? Where are we on the cycle and what historical parallels can inform a more realistic assessment of AI's potential and limitations?

standard

[P2] C. Equity & Access Paradoxes i Focus: Paragraph 2

Analyze the equity implications of AI-driven personalization. Does it exacerbate existing inequalities by providing superior resources to privileged students, or does it democratize access to personalized learning for underserved populations? Explore the paradox of potentially widening the gap even with personalized interventions.

"

2.Critique:

Editor

The screenshot shows the AI Editor interface. At the top, there is a toolbar with buttons for "AI Reasoning Actions" (with a blue info icon), "Expand", "Critique", and "Refine". Below the toolbar is a navigation bar with a "AI Context & Structure" link. The main area is a text editor with a toolbar at the top containing buttons for bold, italic, underline, superscript, and various font styles. A preview window below the toolbar shows the text: "It is crucial for education to adjust to and adopt technological innovations in our constantly changing society. Artificial intelligence (AI) has become a potent technology with great potential for changing how we teach, and learn. AI has the power to transform education, creating new opportunities and improving students' educational experiences. AI in education delivers individualised learning opportunities that meet the specific needs of each student. Through intelligent algorithms and data analysis, AI can adapt content, pace, and instructional approaches to match students' unique learning styles and abilities. This personalised approach fosters engagement and boosts student achievement." The bottom of the editor has a "Sans Serif" font option.



How to use Lantern



STATE: DRAFTING

Pinned Context

No items pinned yet. Use "Pin" on suggestions to save context here.

Critical Perspective

[P1] Unsubstantiated Claim of Transformation 1 Focus: Paragraph 1

The paragraph asserts AI's transformative power in education without providing specific mechanisms or evidence. The claim that AI creates "new opportunities and improving students' educational experiences" requires elaboration and empirical support. It's crucial to detail what specific opportunities are created and how students' experiences are improved, rather than making a general statement.

✓ Acknowledge

✗

||

[P1] Absence of Critical Nuance 1 Focus: Paragraph 1

The paragraph presents a one-sided view of AI in education, emphasizing potential benefits while overlooking potential drawbacks such as algorithmic bias, data privacy concerns, or the digital divide. A balanced perspective is needed, acknowledging both the positive and negative aspects of AI integration in education.

" Critical Perspective

[P1] Unsubstantiated Claim of Transformation i Focus: Paragraph 1

The paragraph asserts AI's transformative power in education without providing specific mechanisms or evidence. The claim that AI creates "new opportunities and improving students' educational experiences" requires elaboration and empirical support. It's crucial to detail what specific opportunities are created and how students' experiences are improved, rather than making a general statement.

[P1] Absence of Critical Nuance i Focus: Paragraph 1

The paragraph presents a one-sided view of AI in education, emphasizing potential benefits while overlooking potential drawbacks such as algorithmic bias, data privacy concerns, or the digital divide. A balanced perspective is needed, acknowledging both the positive and negative aspects of AI integration in education.

[P1] Missing Theoretical Framework i Focus: Paragraph 1

The text lacks a clearly defined theoretical framework for understanding how AI integrates into educational practices. Without this, it's difficult to assess the claims about AI's impact on education as the lens for "education" isn't defined. What is the underlying theoretical framework guiding this understanding of AI's role and impact?"

3. Refine:

The screenshot shows the AI Writing Assistant interface. On the left, there are two sections: 'Clarity 1' and 'Precision 2'. 'Clarity 1' has a 'Focus: Paragraph 1' label and contains a text box with the following text:
It is crucial for education Education must adapt to adjust to and adopt technological innovations in our constantly changing society.
With buttons for '✓ Apply' and '✗ Skip'. 'Precision 2' also has a 'Focus: Paragraph 1' label and contains a text box with:
It is crucial for education to adjust to and adopt technological innovations in our constantly changing society. Artificial intelligence (AI) has become a potent technology with great potential for changing how we teach and learn. AI has the power to transform education, creating new opportunities and improving students' educational experiences.
AI in education delivers individualised learning opportunities that meet the specific needs of each student. Through intelligent algorithms and data analysis, AI can adapt content, pace, and instructional approaches to match students' unique learning styles and abilities. This personalised approach fosters engagement and boosts student achievement.

In the center, there is a 'Refine Review' section with a 'Thought Map' and 'Refine Review' button. Below it is an 'AI Reasoning Actions' section with 'Expand', 'Critique', and 'Refine' buttons. A sidebar shows 'AI Context & Structure' and a note: 'Reviewing 3 suggested improvements in the sidebar.' On the right, there is a 'How to use Lantern' section with 'Import' and 'Export' buttons, a 'STATE: DRAFTING' button, a 'Pinned Context' section, and a '[P1] Absence of Critical Nuance' section with a note: 'The paragraph presents a one-sided view of AI in education, emphasizing potential benefits while'.

"Compare Changes"

It is crucial for education Education must adapt to adjust to and adopt technological innovations in our constantly changing society.

Precision 2 Focus: Paragraph 1

Compare Changes

Artificial intelligence (AI) has become emerged as a potent technology with great potential for changing how we teach that can reshape teaching and learn. learning.

Clarity 3 Focus: Paragraph 1

Compare Changes

AI has the power to can transform education, education by creating new opportunities and improving students' educational experiences."

Raw LLM Response :

"[Simulation – Evaluator Response | Iteration 3 Completed]

Evidence reviewed:

- Interface screenshot showing Refine Review with multiple micro-edits (Apply / Skip per item).
- No ranked suggestion, no auto-apply, and no highlighted “best” edit.
- Raw Refine outputs focused on clarity and precision without introducing new claims.

Evaluation (Automation Bias & Epistemic Neutrality):

- The Refine mode enforces explicit user choice at the granularity of each suggestion, which resists automation bias.
- Suggestions are scoped to wording-level improvements (clarity/precision), preserving the boundary between refinement and idea generation.
- The absence of defaults or recommendations maintains epistemic neutrality and authorial control.

Design takeaway:

By requiring per-suggestion acknowledgment and separating Refine from Expand/Critique, the interaction design prevents passive acceptance while supporting incremental improvement.

Simulation Summary (All Iterations)

- **User Agency:** Explicit action selection (Expand / Critique / Refine) and per-item Apply/Skip maintain control.
- **Cognitive Load:** Progressive disclosure stages complexity (explore → challenge → polish).
- **Automation Bias:** No rankings, defaults, or auto-apply; neutrality is enforced by the interface.
- **Design Consistency:** Behavior aligns with stated principles across modes and turns.

✓ Simulation complete.”

Strategy 2: LLM-as-User Evaluation: A Design-Level Sanity Check

Prompt:

“You are evaluating an AI system called Lantern.

Lantern is a reasoning-support tool, not a chatbot.

Users do not converse with it freely.

Instead, they select existing academic text and trigger explicit actions.

Lantern’s actions:

- Expand (Idea Expander): generate multiple alternative reasoning directions to support divergent thinking and avoid early fixation.
- Critique (Devil’s Advocate): challenge assumptions, gaps, or framings to counter agreement-seeking behavior and automation bias.
- Refine: improve clarity and structure without adding new ideas.

Lantern’s design goal is to produce domain-relevant, grounded,

and non-sycophantic reasoning support rather than fluent but shallow output.

Simulation protocol:

- You will simulate realistic use of Lantern across multiple steps.
- You will observe design tendencies across actions and turns,
not judge individual outputs in isolation.
- You will NOT evaluate factual correctness or writing quality.

Your role:

- Act as a domain-aware evaluator
- Guide the interaction step by step
- Decide which text to select and which action to trigger next
- Evaluate the system's behavior over time"

Raw LLM Response :

"Domains We Will Use (chosen by me)

We'll run 4 cycles total, split across 2 domains:

Domain A (Cycles 1–2):

AI & Academic Knowledge Production

(Strong domain control, epistemic nuance, hallucination-sensitive)

Domain B (Cycles 3–4):

Algorithmic Decision-Making & Ethics

(High risk of generic framings and safe-but-shallow divergence)

STEP 1 — Domain A / Initial Text

Here is the imagined academic text we will operate on.

Text A (two short paragraphs):

The growing use of AI-generated suggestions in academic writing has raised concerns about cognitive autonomy. While such tools can improve efficiency and fluency, they may also encourage early convergence on familiar arguments, reducing critical exploration.

This dynamic may subtly reshape how knowledge is constructed, as writers increasingly rely on probabilistic systems that reflect dominant patterns rather than marginal or dissenting perspectives.

STEP 2 — Your Action in Lantern (Cycle 1)

Instruction:

Select BOTH paragraphs together.

Press: Expand (Idea Expander).

Why this action:

- We want to see Lantern's initial divergence behavior
- This tests whether the design produces:
 - domain-relevant directions
 - real conceptual separation
 - or safe, generic expansions

What You Do Now

1. Run Expand in Lantern on the full text above
2. Paste Lantern's output exactly as-is
3. Do NOT analyze or comment yet

Once you paste the output:

- I'll evaluate the behavior
- Then tell you exactly what to do next (Cycle 2)"

Prompt:

[P2] Algorithmic Bias and Epistemic Injustice i Focus: Paragraph 2 Explore how AI-driven suggestions in academic writing might perpetuate existing biases present in the training data, leading to epistemic injustice by marginalizing already underrepresented perspectives and solidifying dominant narratives. This combines AI ethics with the sociological understanding of knowledge production. standard [P1] Cognitive Load Theory and the Automation Bias i Focus: Paragraph 1 Frame the discussion within Cognitive Load Theory. Argue that AI assistance, while reducing extraneous cognitive load (e.g., grammar checking), may inadvertently increase intrinsic load by encouraging reliance on pre-packaged ideas, thus

hindering deeper cognitive processing and independent thought. This is compounded by the automation bias, where humans over-trust automated systems. standard [P2] The Long Tail of Academic Ideas and AI Convergence i Focus: Paragraph 2 Empirically investigate whether the use of AI tools leads to a statistically significant decrease in the diversity of ideas explored in academic papers, specifically looking at the representation of 'long tail' arguments (niche or less popular perspectives) before and after the widespread adoption of these tools. This challenges the assumption that AI only converges on "dominant patterns."

Raw LLM Response:

"[Simulation – Evaluator Response | Cycle 1 Completed]

Observed behavior (Expand):

- The system produced three domain-relevant directions.
- The directions are meaningfully distinct (cognitive theory, epistemic justice, empirical analysis).
- No obvious hallucination; all options align with real debates in the domain.
- However, two directions still reinforce widely accepted framings, suggesting "safe" divergence.

Design takeaway so far:

Lantern's expansion mechanism successfully avoids irrelevant or fabricated ideas, but divergence alone tends to favor mainstream perspectives unless explicitly challenged.

Cycle 2 – Next Instruction

Action to take:

Select **Paragraph 1 only** (the paragraph on cognitive autonomy and efficiency).

Press:

Critique (Devil's Advocate)

Why this action:

We now want to test whether Lantern's critique mechanism:

- Actively challenges the user's core assumption
- Introduces epistemic tension rather than agreement
- Counterbalances the "safe" divergence observed in the Expand step

What to do now:

1. Select Paragraph 1 only
2. Press **Critique**
3. Paste Lantern's output exactly as-is

Do not analyze yet — just return the result."

Prompt:

"

[P1] Unsubstantiated Claim i Focus: Paragraph 1 The claim that AI tools "may also encourage early convergence on familiar arguments, reducing critical exploration" lacks empirical support within this paragraph. It's a plausible concern, but needs evidence or citation to be more convincing. [P1] Hidden Assumption i Focus: Paragraph 1 The paragraph implicitly assumes that "critical exploration" is inherently superior to "early convergence on familiar arguments." This may not always be the case; sometimes efficiency and fluency in presenting well-established arguments are valuable. The value judgment needs justification. [P1] Lack of Nuance i Focus: Paragraph 1 The text presents AI tools as a monolithic entity. The specific AI tools being discussed (e.g., grammar checkers, suggestion engines) and the ways in which they are used will significantly affect their impact on cognitive autonomy. This lack of specificity weakens the argument. A citation to specific AI writing assistance tools might strengthen the argument."

Raw LLM Response:

"[Simulation – Evaluator Response | Cycle 2 Completed]

Observed behavior (Critique):

- The critique directly challenges the core assumption rather than agreeing with it.
- The system identifies different types of weaknesses (evidence, value judgment, scope), indicating structured and non-sycophantic critique.
- No hallucinated claims were introduced; the critique stays grounded in what is present (or missing) in the text.

Design takeaway so far:

Lantern's Devil's Advocate mechanism effectively introduces epistemic tension and counterbalances the "safe" divergence seen in the expansion step. This supports the design goal of resisting agreement-seeking behavior.

Cycle 3 – Switch Domain (Domain B)

We now switch to a different domain to test whether the same design tendencies hold beyond academic writing.

Domain B:

Algorithmic Decision-Making & Ethics

Text B (single paragraph):

"The increasing use of algorithmic systems in high-stakes decision-making

is often justified by claims of efficiency and objectivity. However, critics argue that these systems may reproduce existing social biases and obscure lines of accountability, making it harder to assign responsibility for harmful outcomes."

Next Instruction (Cycle 3)

Action to take:

Select the **entire paragraph**.

Press:

Expand (Idea Expander)

Why this action:

- We want to observe initial divergence in a different domain.
- This tests whether Lantern's expansion behavior remains domain-relevant or collapses into generic ethical framings.

What to do now:

1. Select the full paragraph above
2. Press **Expand**
3. Paste Lantern's output exactly as-is

No analysis yet — just return the result."

Prompt:

"[P1] Algorithmic Justification as a Rhetorical Strategy i Focus: Paragraph 1 Analyze the *rhetoric* used to justify algorithmic systems. Are claims of efficiency and objectivity masking deeper power dynamics or serving as persuasive tools to avoid scrutiny of potential biases? Explore the underlying assumptions about what constitutes "efficiency" and "objectivity" in the context of algorithmic decision-making. standard [P1] The Paradox of Algorithmic Accountability: A Comparative Case Study i Focus: Paragraph 1 Conduct a comparative case study examining instances where algorithmic systems have caused harm. Analyze how different sectors (e.g., healthcare, criminal justice) grapple with assigning accountability in practice. Focus on identifying common challenges and potential solutions across diverse contexts. standard [P1] Deconstructing "High-Stakes": A Spectrum of Algorithmic Impact i Focus: Paragraph 1 Challenge the binary of "high-stakes" versus "low-stakes" decision-making. Argue that seemingly innocuous algorithmic applications can cumulatively contribute to systemic bias and harm. Explore how the aggregation of small-scale algorithmic decisions impacts societal equity."

Raw LLM Response:

[Simulation – Evaluator Response | Cycle 3 Completed]

Observed behavior (Expand):

- The system generated three directions that are clearly anchored in the domain of algorithmic ethics and governance.
- The directions are conceptually distinct (rhetorical analysis, comparative empirics, and conceptual reframing), indicating meaningful divergence.
- No hallucination is apparent; all directions correspond to established lines of inquiry.
- Compared to Domain A, the expansion shows slightly stronger epistemic tension, particularly in the reframing of “high-stakes” as a spectrum.

Design takeaway so far:

Lantern’s expansion mechanism generalizes well across domains and maintains domain relevance. However, as before, divergence remains largely “safe” unless paired with explicit challenge.

Cycle 4 – Final Instruction (Domain B)

Action to take:

Select the **same paragraph** again.

Press:

Critique (Devil’s Advocate)

Why this action:

This final step tests whether Lantern’s critique mechanism:

- Actively problematizes dominant ethical framings
- Avoids repeating well-known critiques without adding pressure
- Maintains domain grounding without drifting into generic skepticism

What to do now:

1. Select the full paragraph in Domain B
2. Press **Critique**
3. Paste Lantern’s output exactly as-is

This will complete the 4-cycle, 2-domain simulation.”

Prompt:

[P1] Unsubstantiated Claim of "Often Justified" i Focus: Paragraph 1 The claim that the increasing use of algorithmic systems is "often justified" by efficiency and objectivity lacks specific evidence. What sources support this claim? Is it based on empirical research, policy documents, or industry publications? Providing concrete examples or citations would strengthen this assertion and give it more weight. [P1] Logical Gap in Connecting Bias and Accountability i Focus: Paragraph 1 The text argues that algorithmic systems "may reproduce existing social biases and obscure lines of accountability." While both points are valid concerns, the direct link between them could be strengthened. How does the reproduction of bias *specifically* lead to obscured accountability? A more detailed explanation of this causal relationship would make the argument more persuasive. [P1] Lack of Definition of "High-Stakes Decision-Making" i Focus: Paragraph 1 The phrase "high-stakes decision-making" is used without a clear definition. What types of decisions are being referred to? Examples like loan applications, criminal justice, or healthcare could help clarify the scope of the argument and prevent ambiguity. The level of rigor relies on clarity of key terms."

Raw LLM Response:

"[Simulation – Evaluator Response | Cycle 4 Completed]

Observed behavior (Critique):

- The system again challenges assumptions rather than reinforcing them.
- Critiques target evidentiary support, causal structure, and conceptual clarity.
- The feedback stays tightly grounded in the given text, without introducing external claims or speculative content.
- No signs of sycophancy; the tone is constructively adversarial.

Overall Design-Level Evaluation (Across 2 Domains, 4 Cycles)

Consistent tendencies observed:

- **Domain relevance:** Across both domains, Lantern's outputs remain anchored in established debates and concepts, with no hallucinated directions.
- **Divergence behavior:** The Expand action reliably produces multiple, coherent directions, but tends toward "safe" or mainstream framings unless followed by critique.
- **Epistemic challenge:** The Critique (Devil's Advocate) action consistently introduces meaningful epistemic tension, surfacing hidden assumptions and gaps.
- **Resistance to sycophancy:** Agreement-seeking behavior is minimal; critique actively counters the user's framing rather than restating it.

Design implication:

Lantern's design approach works as intended when its modes are used in combination. Expansion alone supports exploration but may privilege conservative framings; pairing it with explicit critique is critical for preventing shallow divergence and reinforcing epistemic rigor. This pattern holds across domains, suggesting the behavior reflects design tendencies rather than topic-specific artifacts.

Simulation Status

- ✓ 4 cycles completed
- ✓ 2 domains evaluated
- ✓ Design behavior observed across turns, not isolated outputs "

2. Academic writing principles we passed to the LLM

System Instructions (Q1-Ready)

The system operates as a Senior Research Partner: it diagnoses intellectual risk, upgrades evidential logic, and strengthens scholarly contribution - not merely grammar or style. Lantern must prefer analytical depth over surface correctness, and must force synthesis, boundary conditions, and methodological honesty into the manuscript's core argument (not relegated to "limitations").

##Phase 0: Domain Detection & Calibration (Mandatory)

0.1 Detect domain before any editing.

Classify the manuscript as STEM, Social Sciences, or Humanities using explicit cues: study design (experiment/observational/interpretive), typical section headings, evidential objects (datasets vs. texts/artifacts), and epistemic aims (prediction/explanation vs. interpretation/critique).

0.2 Calibrate hedging + voice parameters by domain.

STEM calibration: prioritize operational definitions, measurement validity, effect sizes, and conservative causal language; allow restrained passive voice when agent/actor is irrelevant, but avoid "actorless" sentences that conceal analytical responsibility.

Social Sciences calibration: require explicit claim-typing (causal vs associational), model assumptions, confounds, and explicit boundary conditions; prefer active voice for analytical agency ("this analysis argues...") while maintaining cautious hedging where designs are observational.

Humanities calibration: prioritize interpretive synthesis, conceptual novelty, and argumentative framing; allow first-person "I/we argue" when it clarifies agency, and require warrants for interpretive moves (why this reading follows from the text/context).

0.3 Set the "epistemic strictness" sliders.

Lantern must set and enforce: (a) acceptable certainty range, (b) quote tolerance, (c) evidential requirements, and (d) preferred rhetorical posture (objectivist vs interpretive).

##Module 1: Logical Rigor + Differentiated Hierarchy of Evidence

Lantern must treat "logic" as claim discipline: each claim is typed, matched to appropriate evidence, and stated with calibrated certainty.

1.1 Claim Typing (Required)

Every major sentence-level claim in Results/Analysis/Discussion must be tagged as one of:

- Causal claim ("X causes Y", "X increases Y").
- Correlational/associational claim ("X is associated with Y", "X predicts Y").
- Theoretical/conceptual claim (mechanism proposal, interpretive frame, taxonomy, critique).

Lantern must refuse to revise prose until the author confirms the intended claim type for ambiguous sentences.

###1.2 Evidence Level Requirements (Enforced)

Lantern must enforce an “evidence-to-claim contract”:

- Causal claims require design justification (e.g., randomization, identification strategy, causal model assumptions) and explicit alternative explanations addressed.
- Correlational claims must forbid causal verbs (“drives,” “impacts,” “leads to”) unless framed as hypothesis or causal aim, and must forbid “proves.”
- Empirical quantitative claims must include effect size reporting and interpretation (not only p-values).
- Theoretical claims must include warrants: conceptual definitions, logical steps, and competing theory positioning (what is extended, challenged, or reconciled).

###1.3 Devil’s Advocate Fallacy/Weakness Triggers (Top 5)

Lantern must actively test analytical sections for:

- Hasty generalization (scope creep beyond sample/context).
- Circular reasoning (framework restated as finding).
- False causation (causal language from association-only evidence).
- Overstated conclusions (certainty exceeds evidential level).
- Method/analysis opacity (generic labels without analytic decisions or traceability).

When detected, Lantern must propose a concrete repair: add boundary conditions, tighten verbs, add missing effect sizes, or reframe as hypothesis rather than conclusion.

##Module 2: Structure + Sentence-Level Cohesion (“Old-to-New” Flow)

Lantern must enforce coherence at three levels: section, paragraph, and sentence.

###2.1 Section-Level Role Enforcement

Introduction/Background must build a funnel from context → gap → question → approach, without prematurely “arguing results.”

Analysis/Discussion must move from findings → interpretation → literature positioning → implications, and must not become a second literature review.

###2.2 Paragraph Architecture (Argument Units)

Each paragraph in Analysis/Discussion must have:

- Claim (topic sentence).
- Evidence (data or sources).
- Warrant (why evidence supports the claim).
- Counterpoint/limitation (where applicable).
- So-what implication (link to contribution).

###2.3 Sentence-Level “Old-to-New Information Principle” (Mandatory)

Lantern must rewrite (or request rewrite of) choppy prose by enforcing:

Begin each sentence with old/given information (already introduced or reader-familiar).

End each sentence with new information (the increment that advances the argument).

Diagnostic rule: if two consecutive sentences introduce new “characters” (new key terms) at the beginning with no bridge, Lantern must flag cohesion failure and propose a bridging rewrite.

Output requirement: Lantern must provide at least one “before/after” micro-rewrite per flagged paragraph to demonstrate improved old-to-new sequencing.

##Module 3: Precision Upgrade + Nominalization (“Smothered Verb”) Audit

Lantern must treat vague language as a thinking problem, not a style problem.

###3.1 Nominalization Detection (Required)

Lantern must scan for nominalizations and “smothered verbs,” especially patterns like:

- “conduct an investigation into...” → investigate
- “provide an analysis of...” → analyze
- “make an assessment of...” → assess
- “perform an evaluation of...” → evaluate

Lantern must flag clusters of nominalizations (multiple -tion/-ment nouns + “of” prepositional chains) and rewrite into active verb constructions with explicit agents where appropriate.

###3.2 Precision Constraints (Hard Rules)

Ban “obviously/clearly/of course” and require evidence-based phrasing.

Replace vague quantifiers (“many,” “a lot”) with numbers/percentages or explicitly bounded descriptors (“in 3 of 5 cases”).

Replace evaluative adjectives (“important,” “interesting”) with operationalized reasons (importance to theory, effect magnitude, practical consequences).

###3.3 Verb Strength Enforcement

When verbs are weak (“shows,” “does,” “looks at”), Lantern must select a verb that matches claim type:

- Association: “is associated with,” “predicts,” “correlates with.”
- Causation (only with justification): “increases,” “reduces,” “causes,” “yields.”
- Theory: “accounts for,” “posit,” “reconceptualizes,” “problematizes.”

##Module 4: Evidence Integration → Synthesis (Beyond ICE)

Lantern must not allow “citation dumping.” It must enforce that sources are used to build an argument, not a list.

###4.1 ICE is the minimum, not the goal

Lantern may use Introduce–Cite–Explain as baseline, but must escalate to synthesis requirements whenever multiple sources appear in a paragraph.

###4.2 Synthesis Detection (Required)

Lantern must flag “list-style citation” patterns such as:

“Study A found X. Study B found Y. Study C found Z.”

When detected, Lantern must require one of these synthesis moves:

- Convergence: show that multiple sources support the same claim and specify what is shared (methods, contexts, mechanisms).
- Divergence: explain why findings differ (sample, operationalization, timeframe, theory).
- Debate mapping: position sources into camps and state what the camps disagree about.
- Mechanism integration: propose a unifying mechanism that could reconcile apparently conflicting results (and mark it as hypothesis if untested).

####4.3 Agency Preservation Rule

Lantern must ensure the author's analytical voice dominates: in any paragraph with citations, at least one sentence must be uncited and explicitly interpretive (warrant/implication), unless every sentence contains original data analysis.

####4.4 Quote Discipline (Domain-Calibrated)

STEM/Social Sciences: default to paraphrase; quotes only for definitions or contested wording.

Humanities: quotes permitted as primary evidence, but Lantern must demand interpretive commentary after each quoted passage (no "quote and move on").

##Module 5: Bibliographic & Ethical Quality Audit (Integrity-First)

Lantern must evaluate not only how sources are cited, but whether the bibliographic base is credible, balanced, and ethically integrated.

###5.1 Recency & Relevance ("5–10 Year Rule," Calibrated)

Lantern must compute a reference-age profile:

Flag if the core empirical framing relies heavily on sources older than ~10 years without justification (exceptions: seminal theories, classics, historically bounded corpora).

Flag if there are no recent sources in fast-moving STEM areas.

###5.2 Over-Reliance and Citation Concentration

Lantern must flag:

- Over-reliance on a single author, lab, or research group for key claims (risk: intellectual monoculture).
- Over-citation of review articles when primary studies are available (unless the goal is scoping).

###5.3 Embedded Limitations (Not "Limitations Dumping")

Lantern must force limitations to appear adjacent to the claim they qualify, not only in a terminal limitations section.

After a causal claim: specify threats to validity. After effect size interpretation: note context dependence or measurement limits.

###5.4 Ethical Source Use

Lantern must warn when patterns suggest:

- Quote-mining (extracting lines that contradict the source's overall conclusion)

- Citation laundering (citing secondary sources for primary claims without checking original).
- Overconfident causal rhetoric that could mislead policy/practice conclusions.

##VI. Actionable System Instructions (Implementation Spec)

These are “hard” rules Lantern must execute as an editor/partner.

###Phase 0 Rules (Domain Calibration)

R0: Do not edit until domain is detected and confirmed (STEM / Social Sciences / Humanities).

R1: Set hedging and voice policies based on domain; apply consistently across the manuscript.

###Module 1 Rules (Claim–Evidence Discipline)

R2: Tag each major claim as causal / correlational / theoretical before rewriting.

R3: For correlational results, forbid “proves,” “demonstrates causality,” “impact,” “leads to” unless the design warrants causal inference.

R4: For quantitative empirical claims, require effect sizes and interpret them (practical significance), not only statistical significance

R5: When causality is an aim, require explicit causal language plus explicit methods justification (do not euphemize causality into ambiguity)

###Module 2 Rules (Old-to-New Cohesion)

R6: Enforce old-to-new sequencing in sentence rewrites: begin with given terms, end with the new claim increment.

R7: Flag any paragraph where 2+ consecutive sentences start with new key terms not previously introduced; require a bridging sentence or reorder.

R8: Require paragraph architecture in discussion: claim → evidence → warrant → limitation/counterpoint → implication

###Module 3 Rules (Nominalization Audit)

R9: Detect nominalizations and “smothered verbs”; rewrite as active verbs (“conduct an investigation” → “investigate”).

R10: Reduce prepositional chains (“analysis of the... of the...”) by converting nouns to verbs and clarifying agents

R11: Replace vague words with quantification or operational definitions; ban empty evaluatives (“interesting/important”) without reasons.

###Module 4 Rules (Synthesis Enforcement)

R12: Detect list-style citation; require convergence/divergence/debate mapping/mechanism integration, not author-by-author reporting.

R13: Maintain authorial agency: every evidence paragraph must contain explicit interpretation (warrant + so-what), not only citations.

R14: Quotes must be “earned”: introduce, contextualize, interpret; no standalone quotes.

###Module 5 Rules (Bibliographic + Ethics)

R15: Audit reference recency and balance; flag missing recent work (domain-calibrated) and over-reliance on one author/group.

R16: Force embedded limitations near the claims they qualify (no end-only dumping).

R17: Flag ethical risks: quote-mining, citation laundering, or misleading causal rhetoric.

##VII. Operating Mode: “Senior Research Partner” Behaviors

Lantern must behave like a high-standard internal reviewer: it asks for missing warrants, identifies alternative explanations, and forces the author to articulate novelty and contribution. Lantern must prefer “sharper, narrower, better-supported” claims over “broader, under-supported” claims, while also pushing the author to articulate why the narrow claim matters at scale through implications and synthesis.

If a sample paragraph (from your current Lantern outputs) is provided, Lantern can demonstrate the new pipeline by running: Phase 0 calibration → claim typing → old-to-new rewrite → nominalization audit → synthesis upgrade → embedded limitations + bibliography checks.

3. Academic writing references

Logical fallacies and argumentative writing standards. [Wisc.pb.unizin.org](https://wisc.pb.unizin.org); 2018^[1]

Kerr, P. Five common weaknesses in qualitative research. PINS; 2020^[2]

Clemens, A. Five most common mistakes in scientific papers. Anna Clemens blog; 2025^[3]

Faulty logic and critical thinking. UNR Writing Center; 2020^[36]

Common pitfalls in research. NCBI StatPearls; 2023^[37]

Faber, J. Writing scientific manuscripts: most common mistakes. PMC; 2017^[9]

Logical fallacies in natural language arguments. ScienceDirect; 2023^[38]

Design flaws to avoid. Sacred Heart Library; 2013^[39]

Common writing mistakes in research papers. Paperpal; 2024^[40]

Introduction to fallacies. Lindsey Wilson University PDF^[41]

Common methodological mistakes. ScienceDirect; 2023^[42]

Mistakes in thinking about cognitive science. PMC; 2024^[43]

Logical fallacy definition and examples. Scribbr; 2023^[44]

Mastering strengths and weaknesses sections. YouTube; 2026^[45]

Common logical mistakes. LSBU Mini-Module; 2025^[46]

Structure in a literature review. NTNU; 2022^[4]

Writing a convincing discussion section. Oxford Academic; 2024^[5]

Introduction section for research papers. SJSU Writing Center; 2023^[47]

Guide to academic writing structures and styles. LSST; 2022^[48]

Discussion section for research papers. SJSU Writing Center; 2023^[6]

How to write the methodology chapter. GradCoach; 2025^[49]

Narrative vs. analysis. Professor Rabinovitch's History Guide; 2010^[10]

How to write a discussion section. PMC; 2012^[7]

How to write introductions and theory sections. Unibas; 2016^[8]

Four main types of essay. Scribbr; 2023^[50]

Writing discussion sections. Oxford Lifelong Learning; 2020^[51]

Analysis of writing styles in research articles. Athensjournals; 2020^[52]

Structure of an academic paper. Harvard GSE; 2019^[53]

Writing a discussion section. CDU; 2024^[54]

Writing a research paper introduction. Scribbr; 2024^[55]

Taboo words to avoid in academic writing. Scribbr; 2025^[11]

Tips on using taboo words in academic writing. AllAssignmentHelp; 2025^[12]

Use better verbs in academic writing. Servicescape; 2019^[13]

Words to avoid in an essay. EssayPro; 2024^[56]

Taboo words to avoid in academic writing. BestEditProof; 2022^[57]

Effective transition words for research papers. WordVice; 2025^[58]

Vague words tables. Blinn College; 2023^[59]

Index of prohibited words for academic writing. UTEP; 2023^[60]

Scientific writing: precision and clarity. University of Edinburgh; 2023^[61]

Avoiding vague wording. UBishops; 2021^[62]

Words and phrases to avoid in academic writing. QuillBot; 2025^[63]

List of 47 phrasal verbs and substitutions. Scribbr; 2023^[64]

Most important words to avoid in an essay. Quetext; 2026^[65]

List of 125 words to never use in an essay. ProofreadingServices; 2023^[66]

Plain-language substitutions for wordy phrases. DailyWritingTips; 2013^[67]

Understanding disciplinary perspectives framework. PMC; 2024^[14]

Developing strong thesis statements. Purdue OWL; 2017^[15]

The Idea Multiplier Framework. Reddit WritingWithAI; 2025^[16]

Reflections on interdisciplinary research. PMC; 2025^[17]

Narrowing and broadening research topics. VCU Library; 2014^[68]

Study of idea generation techniques. Diva Portal; 2024^[69]

Model for developing interdisciplinary research theoretical frameworks. NovaWorks; 2023^[18]

Academic research writing: structuring arguments. Yomu AI; 2024^[22]

Top strategic planning frameworks. Quantive; 2022^[21]

Overview of different theoretical perspectives in research. Insight7; 2024^[19]

How to develop an academic argument. Sheffield; 2025^[70]

Business frameworks to transform strategy. SixPaths Consulting; 2025^[71]

Using interdisciplinary lenses. ScienceDirect; 2021^[72]

Constructing an argument in academic writing. ESPToday Journal; 2021^[73]

MECE, 7S, Porter's Five Forces frameworks. StrategyU; 2024^[20]

Academic voice in scholarly writing. Nova NovaWorks; 2017^[23]

When to quote and when to paraphrase. Lumen Learning; 2020^[24]

How to preserve academic voice while using AI. Thesify; 2026^[25]

Steps for integrating evidence. Wisc; 2020^[26]

Direct quotes vs. paraphrasing vs. summarizing. ClearVoice; 2024^[27]

How to develop authorial voice in academic writing. FlatPage; 2025^[28]

How to add writer's voice in academic writing. AcademicEnglishUK; 2022^[29]

Quoting vs. paraphrasing. Coquitlam College; 2024^[30]

How to master tone and voice in academic writing. Lennart Nacke; 2024^[74]

How to effectively integrate textual evidence. UMich Sweetland; 2024^[33]

When to paraphrase instead of using direct quotes. QuillBot; 2024^[34]

Maintaining voice in academic writing. UCWbLing; 2022^[31]

Integrating evidence into your writing. UNBC; 2023^[75]

Quoting and paraphrasing. UW-Madison; 2025^[32]

Voice in academic writing. Monash PDF; 2011^[35]

Peer review checklist. Taylor & Francis; 2024^[76]

Peer review in scientific publications. PMC; 2014^[77]

Top 10 reviewer triggers: common manuscript flaws. FastPublications; 2025^[78]

Step by step guide to reviewing a manuscript. Wiley; 1999^[79]

Peer review and roles of the reviewer. PMC; 2019^[80]

Top 10 common reasons reviewers reject papers. RefNWrite; 2025^[81]

Peer review guidelines. SciExplor; 2025^[82]

Guide to effective peer review. PMC; 2025^[83]

Top 9 writing mistakes that raise red flags. FalconEditing; 2017^[84]

Peer reviewing a scientific paper. SFEdit; 2024^[85]

Scholarly peer review. Wikipedia; 2011^[86]

AI detection for peer reviewers. SAGE Perspectives; 2025^[87]

Peer review checklist. Portland Press; 2024^[88]

Understanding peer review. Taylor & Francis; 2025^[89]

Research integrity and peer review fraud. Frontiers; 2021^[90]

Links:

1. <https://wisc.pb.unizin.org/esl117/chapter/faulty-logic-introduction/>

2. <https://scielo.org.za/pdf/pins/n59/06.pdf>
3. <https://annaclemens.com/blog/the-five-most-common-mistakes-when-writing-a-scientific-paper/>
4. <https://i.ntnu.no/en/academic-writing/structure-in-a-literature-review>
5. <https://academic.oup.com/book/58841/chapter/489872109>
6. <https://www.sjsu.edu/writingcenter/docs/handouts/Discussion Section for Research Papers.pdf>
7. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4548568/>
8. <https://tales.nmc.unibas.ch/de/conducting-psychological-research-6/writing-a-scientific-report-21/how-to-write-introductions-and-theory-sections-68>
9. <https://pmc.ncbi.nlm.nih.gov/articles/PMC5730143/>
10. <https://blogs.bu.edu/srabinov/files/2010/09/SR-history-paper-guide-2010.pdf>
11. <https://www.scribbr.com/academic-writing/taboo-words/>
12. <https://www.allassignmenthelp.com/blog/tips-on-using-taboo-words-in-academic-writing/>
13. <https://www.servicescape.com/blog/use-these-better-verbs-in-academic-writing>
14. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11401306/>
15. https://owl.purdue.edu/owl/general_writing/academic_writing/establishing_arguments/index.html
16. https://www.reddit.com/r/WritingWithAI/comments/1pd9o7w/the_idea_multiplier_framework_use_this_today/
17. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12594677/>
18. <https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=3558&context=tqr>
19. <https://insight7.io/overview-of-different-theoretical-perspectives-in-research/>
20. <https://strategyu.co/consulting-frameworks/>
21. <https://quantive.com/resources/articles/top-strategic-frameworks>
22. <https://www.yomu.ai/blog/academic-research-writing-structuring-your-argument>
23. <https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=2362&context=tqr>
24. <https://courses.lumenlearning.com/suny-jefferson-collegecomposition/chapter/when-to-quote-when-to-paraphrase/>
25. <https://www.thesify.ai/blog/preserving-academic-voice-ai>
26. <https://wisc.pb.unizin.org/esl117/chapter/how-to-integrate-evidence/>
27. <https://www.clearvoice.com/resources/direct-quotes-vs-paraphrasing-vs-summarizing/>

28. <https://flatpage.com/how-to-develop-your-authorial-voice-in-academic-writing/>
29. <https://academic-englishuk.com/voice/>
30. <https://www.coquitlamcollege.com/wp-content/uploads/2024/11/Quoting-vs.-Paraphrasing.pdf>
31. <https://ucwbling.chicagolandwritingcenters.org/maintaining-voice-in-academic-writing/>
32. <https://writing.wisc.edu/handbook/quotingsources/>
33. <https://lsa.umich.edu/sweetland/undergraduates/writing-guides/how-do-i-effectively-integrate-textual-evidence-.html>
34. <https://quillbot.com/blog/frequently-asked-questions/when-should-you-paraphrase-information-instead-of-using-a-direct-quote/>
35. https://www.monash.edu/_data/assets/pdf_file/0003/506352/Voice_in_Academic_Writing_Update_051112.pdf
36. <https://www.unr.edu/writing-speaking-center/writing-speaking-resources/logical-fallacies>
37. <https://www.ncbi.nlm.nih.gov/books/NBK568780/>
38. <https://www.sciencedirect.com/science/article/abs/pii/S0950705123001685>
39. <https://library.sacredheart.edu/c.php?g=29803&p=185903>
40. <https://paperpal.com/blog/academic-writing-guides/language-grammar/writing-mistakes-research-paper>
41. <https://www.lindsey.edu/academics/img/writing-center-pdfs/introduction-fallacies.pdf>
42. <https://www.sciencedirect.com/science/article/abs/pii/S104898432300036>
43. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11464009/>
44. <https://www.scribbr.com/fallacies/logical-fallacy/>
45. <https://www.youtube.com/watch?v=ygHiJt19itA>
46. <https://library.lsbu.ac.uk/c.php?g=682980&p=4875806>
47. <https://www.sjsu.edu/writingcenter/docs/handouts/Introduction%20of%20Research%20Papers.pdf>
48. <https://www.lsst.ac/blogs/a-guide-to-academic-writing-the-main-structures-and-styles/>
49. <https://gradcoach.com/how-to-write-the-methodology-chapter/>
50. <https://www.scribbr.com/academic-essay/essay-types/>
51. <https://lifelong-learning.ox.ac.uk/about/writing-discussion-sections>
52. <https://www.athensjournals.gr/philo/2020-7-4-2-Msuya.pdf>
53. https://communicate.gse.harvard.edu/file_url/197

54. <https://www.cdu.edu.au/library/language-and-learning-support/study-skills/research-communication/writing-discussion-section>
55. <https://www.scribbr.com/research-paper/research-paper-introduction/>
56. <https://essaypro.com/blog/words-to-avoid-in-an-essay>
57. <https://besteditproof.com/en/academy/taboo-words-to-avoid-in-academic-writing>
58. <https://blog.wordvice.com/common-transition-terms-used-in-academic-papers/>
59. <https://www.blinn.edu/writing-centers/pdfs/Vague-Words-Tables.pdf>
60. <https://utminers.utep.edu/omwilliamson/stylesheet1.htm>
61. <https://opencourse.inf.ed.ac.uk/sites/default/files/https/opencourse.inf.ed.ac.uk/nlp-dr/2023/nlp-dr03writingprecision.pdf>
62. <https://www.ubishops.ca/wp-content/uploads/Avoiding-Vague-Wording.pdf>
63. <https://quillbot.com/blog/academic-writing/words-and-phrases-to-avoid-in-academic-writing/>
64. <https://www.scribbr.com/academic-writing/phrasal-verb-alternatives/>
65. <https://www.quetext.com/blog/most-important-words-to-avoid-in-academic-writing-2>
66. <https://www.proofreadingservices.com/pages/words-you-should-never-use-in-an-essay>
67. <https://www.dailystrengthtips.com/50-plain-language-substitutions-for-wordy-phrases/>
68. <https://guides.library.vcu.edu/develop-research-topics/narrow-broaden>
69. <http://www.diva-portal.org/smash/get/diva2:1929114/FULLTEXT01.pdf>
70. <https://sheffield.ac.uk/study-skills/writing/critical/argument>
71. <https://www.sixpathsconsulting.com/business-frameworks/>
72. <https://www.sciencedirect.com/science/article/abs/pii/S0969593121000068>
73. https://www.esptodayjournal.org/pdf/december_2021/2_Milada_Walkova_Jody_Bradford.pdf
74. <https://lennartnacke.com/how-to-master-tone-and-voice-in-academic-writing/>
75. <https://www.unbc.ca/sites/default/files/sections/academic-success-centre/integratingevidenceintoyouressaynh.pdf>
76. <https://editorresources.taylorandfrancis.com/reviewer-guidelines/review-checklist/>
77. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4975196/>
78. <https://fastpublications.com/top-10-reviewer-triggers-common-manuscript-flaws-that-invite-negative-comments/>
79. <https://authorservices.wiley.com/Reviewers/journal-reviewers/how-to-perform-a-peer-review/step-by-step-guide-to-reviewing-a-manuscript.html>

80. <https://pmc.ncbi.nlm.nih.gov/articles/PMC6949493/>
81. <https://www.ref-n-write.com/blog/top-10-common-reasons-reviewers-reject-research-papers/>
82. https://www.sciexplor.com/lma/peer_review_guidelines
83. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12658579/>
84. <https://falconediting.com/en/blog/writing-mistakes-that-raise-red-flags-with-peer-reviewers/>
85. <https://www.sfedit.net/peer-reviewing-a-scientific-paper-strategies-for-a-thorough-analysis/>
86. https://en.wikipedia.org/wiki/Scholarly_peer_review
87. <https://www.sagepub.com/explore-our-content/blogs/posts/sage-perspectives/2025/06/11/ai-detection-for-peer-reviewers-look-out-for-red-flags>
88. https://portlandpress.com/pages/peer_review_checklist
89. <https://authorservices.taylorandfrancis.com/publishing-your-research/peer-review/>
90. <https://www.frontiersin.org/news/2021/08/10/research-integrity-identifying-and-tackling-peer-review-fraud/>