

Literature Research with AI

From Survey to Synthesis

Scientific Inquiry with AI

Lesson Overview

Created by LLNL Summer 2026 STEM Education Research Team

Student Team: Ramina Amino, Jahanvi Chamria, Arya Ferozy, Bryanna Gonzalez, Tai Le, Zedikiah McAdams, John Navarra, Joshua Sarabia, Abdurrahman Raza

Faculty Team: Praveen Pathak, David Rakestraw, David Strubbe, Brian Utter

Purpose and Scope

This document presents the Literature Research with AI lesson in the course's standard three-phase structure: Pre-Class Preparation, In-Class Investigation, and Post-Class Work. It is the course-facing summary — the version to use for program planning, scheduling, accreditation documentation, and orienting a colleague to what the lesson does.

This is a research-skills lesson rather than an experimental investigation, so detailed structure will be adjusted for the topic. Everything else about the course's structure holds — students prepare individually, investigate in parallel as a class with instructor coaching, and consolidate afterward.

Full instructor detail — minute-by-minute protocol, prompt templates, video specification, assessment rubric, and the evidence base for the design — is in the companion Instructor Guide. The Student Guide is the student-facing version of this lesson, and is the authoritative source for timing, sequence, prompts, and deliverables; this overview is aligned to it.

Lesson Summary

Each student conducts a literature review on a research question of their own through a five-step AI-assisted workflow: claim survey, citation mapping, synthesis matrix, grounded synthesis, and deep research. Rather than run a physical experiment, students learn to survey a field, verify AI-generated claims against primary sources, and identify a genuine gap or contested point as the first step of original inquiry.

The lesson is built on manual-first before using AI approach. Before class, students perform a keyword literature search by hand in Google Scholar and read one paper without AI, generating a felt baseline — “the friction” — against which every AI-assisted result is later judged. In class, they run the full workflow on their own question, with delegation to the AI increasing at each step and a required verification move at the end. The in-class investigation runs across two class periods and closes with a segment on learning from a corpus the student has personally verified, followed by a short reflection. The major written deliverable, including the gap statement, is produced after class.

Lesson at a Glance

Audience	First-year STEM majors; no prior research experience assumed. For most students this is their first structured encounter with literature review as a discipline.
Time	~2 hours pre-class (manual literature research 50 min • demonstration video ~20 min including pauses • Socratic tutor session ~20 min • written summary 10 min • tool-access check 5 min • student-guide review 10 min); 220 min in class, across two class periods (Activation 15 min • five-step workflow 165 min • NotebookLM exploration 30 min • Reflect 10 min); ~2.5 hours post-class (review and step-comparison table 30 min • Literature Research Summary Document 90 min • retrieval practice 30 min).
Structure	Pre-Class Preparation → In-Class Investigation → Post-Class Work. Manual-first literature work and the five-step demonstration (video plus Socratic tutor) happen pre-class. Class time — two 110-minute periods — is devoted to students running the full workflow on their own research questions; the instructor sets the break point between the two meetings according to the pace of the class. Generation of a workflow summary table and an expanded literature research summary are prepared after class.
Materials	Demonstration video (link distributed with the prelab); Prelab Tutor Document (companion file, uploaded by each student to their AI); instructor PowerPoint deck to pace the steps; a general-purpose AI model with web search, file upload, a deep research mode, and the ability to generate files — the Step 2 and Step 3 prompts request a PDF report and an Excel spreadsheet plus an interactive HTML view (ChatGPT is the assumed baseline); spreadsheet software to open the matrix; Google Scholar; ResearchRabbit (Connected Papers or Litmaps as substitutes); NotebookLM (free Google account); institutional library proxy credentials; student whiteboards and markers; projector. No laboratory equipment or safety requirements.
Deliverables	Pre-class: research question, manual search notes including the friction, three-sentence deep-read summary plus one unanswered question, PDFs of 2–3 central papers, two written video pause notes, tutor chat log, written summary of the five steps. In class: refined question, first synthesis matrix (five sources), the student's own written read of that matrix, spot-check record, three closing one-line answers. Post-class: step-comparison table (PDF), Literature Research Summary Document (five elements, 6–10 paper matrix), NotebookLM notebook or link.

Learning Goals

Learning goals are grouped into the course's three standard categories:

- **Phenomena:** the observable pattern that motivates the inquiry — here, the structure of a research literature.
- **Experimental Practices:** the specific analytical skills students apply.
- **Habits of Mind:** the reasoning, communication, and reflection practices students develop.

By the end of this lesson, students will be able to:

Phenomena

- Read a field rather than a paper: identify what a research community treats as settled, what it treats as contested, and the signals that distinguish the two.

Experimental Practices

- Formulate a focused, researchable question by hand, then use AI as a critic to refine it — and explain why question quality determines search quality.
- Execute an AI-assisted research workflow — claim survey, citation mapping, synthesis matrix, and grounded synthesis — using structured, reusable prompts.
- Design the dimensions of a synthesis matrix appropriate to their own research question and use it to evaluate and compare sources.
- Use the integrated deep research tools native to most foundation models, and recognize how those tools combine the workflow they performed step by step.
- Use a curated corpus — their verified papers, plus the deep research report as an unverified reference — in a grounded-Q&A tool such as NotebookLM to deepen their own understanding.

Habits of Mind

- Verify AI-generated claims against primary sources and articulate why this is the same discipline scientists have always applied to any intermediary source.
- Identify gaps, contested points, or extensions not found in the current literature — the first step of original inquiry.
- Recognize when a fully automated “deep research” tool is the right way to get oriented in an unfamiliar topic, and when its output still needs further investigation to meet their objectives.

The Core Idea

Three ideas carry the lesson. They appear in the demonstration video, are rehearsed with the Socratic tutor, and are restated in class — deliberately in the same language each time.

- **Key Concept 1 — AI output sits on a spectrum of trust.** Search tools hand back real links to real papers a student can open. A language model additionally writes summaries, and those must be checked for accuracy before being relied on. Knowing which kind of output you are holding tells you what kind of checking it needs.
- **Key Concept 2 — As the AI does more, the researcher's job shifts from doing to directing and checking.** Across the five steps, delegation grows. The decisions that never transfer are what question to ask, what claims to trust, and what gap to pursue.
- **Key Concept 3 — A synthesis matrix is a discovery tool, not just a summary.** One small claim per cell makes the whole landscape visible at once: agreement, disagreement, and — in the empty cells — knowledge gaps. A cell is also a small, checkable claim, which makes the matrix a verification-friendly format.

The driving question

“Does using AI tools for problem-solving improve undergraduate STEM students’ learning and long-term retention, or does it create a “cognitive offloading” effect that reduces skill-building – and does the outcome depend on how the AI is used, such as answer-giving versus Socratic tutoring?”

Our working hypothesis is: AI may improve learning and retention when used as a Socratic tutor but may reduce skill-building when used only to get direct answers.

This is the shared demonstration question used in the video and the tutor; every student then runs the same workflow on a question of their own. Its general form: *for a question I actually care about, what does the literature treat as settled, what is still contested, what has no one asked — and how much of finding that out can I hand to an AI without losing the ability to tell whether it is right?*

The Five-Step Workflow

The workflow is the backbone of the lesson and is reproduced in the Student Guide. The ordering principle is the amount of work handed to the AI; the obligation in the final column never goes away.

Step	What the student does	What the AI does	What the student must check
1. Claim survey	Frame the question and the hypothesis to be tested	Searches the literature and returns papers with their findings	Whether each summarized finding matches the paper
2. Citation mapping	Choose a seed paper and read the structure around it	Maps what the paper builds on and what has built on it	Whether the stated relationships are real
3. Synthesis matrix	Decide which columns the question requires	Fills one row per paper, leaving unknowns blank	Every cell the student intends to rely on
4. Grounded synthesis	Read the matrix and form an independent view first	Synthesizes agreement, disagreement, and gaps	Whether its synthesis matches what the student saw
5. Deep research	Pose the question, then judge what comes back	Runs the whole workflow automatically	The claims and sources the student intends to use

Part 1 — Pre-Class Preparation

~2 hours

Students complete all pre-class activities individually, though peer discussion is encouraged. The prelab brings every student — regardless of prior research experience — to a common readiness state: a research question of their own, a felt baseline of what manual literature work requires, a

complete viewing of the five-step AI supported literature research process to be replicated in class, and a tutor-checked grasp of the lesson's core ideas.

A. Topic Selection and Manual Literature Research (~50 min)

No AI in this component. This is the lesson's manual-first foundation.

- **Define a research question (10 min).** Each student picks a topic they genuinely care about — or adopts and adapts one from a short list in the Student Guide — and manually composes a focused, researchable question, using the strong-question heuristic: a system or population, a mechanism or variable, and a comparison. The starter list exists so that no student finishes the prelab with a blank page; the starter questions are deliberately underdeveloped so work remains for the students in developing their research questions.
- **Manual search (20 min).** Google Scholar only — pure keyword search. Students must define the key search words they think will bring forward the best sources. Students find the five most relevant-seeming papers; read titles and abstracts; take notes on the friction: whether the search terms captured the question, what was hard to find, what was wasted effort, and what relevant work probably exists under terms they did not think to try.
- **Manual read (15 min).** Read one paper without using AI. Write a three-sentence summary in the student's own words plus one question the paper did not answer.
- **Obtain primary sources (5 min).** Save full-text PDFs of the two or three most central papers. These seed the synthesis matrix students will build in class. If a paper resists access after a genuine attempt, the student documents what they tried and moves on; an abstract is an acceptable fallback, marked as such in the matrix.

The Friction of the Manual Process: The student notes about “friction” in the manual process are the reference experience against which every AI-assisted result in this lesson is judged, and class opens by collecting them. Students need to feel the bottleneck before being shown the tool that relieves it — otherwise the five steps register as features to memorize rather than solutions to problems they have personally had. This is the same logic as teaching manual computation before handing over the calculator: students may never search this way again, but they will spend their careers evaluating tool output, and evaluation requires having once done the thing by hand.

B. Demonstration Video with Pause-and-Predict (~20 min including pauses)

Students watch a video demonstrating all five steps of the AI supported literature review process on the shared demonstration question. The five-step demonstration is delivered by video rather than by lecture so that class time goes for students running the workflow themselves. A video also guarantees every student sees one complete, well-executed run through each step, and any step can be rewatched mid-task.

The video contains two mandatory written pause prompts:

- **Pause 1, before Step 1 is revealed:** write the exact instructions you would give a human research assistant surveying this literature, including what to find, what to extract, and how to handle uncertainty.
- **Pause 2, before Step 4 runs:** write one claim you would bet is settled and one you would bet is contested, and what signal would tell you the difference.

Students bring both written notes to class. The tutor asks about them, the student's hypothesis feeds directly into the Step 1 prompt, and the opening retrieval exercise revisits them.

C. Socratic AI Tutoring Session (~20 min)

Students upload the customized Prelab Tutor Document to their AI foundation model. The tutor reviews the video's core ideas Socratically — the trust spectrum, the delegation gradient and what deep research does and does not replace, verification as standard source discipline, and ownership of the question — drawing on the student's pause notes as diagnostic input. The session closes with an engagement score (a personal target, not a grade) and a conceptual roadmap. Students share the chat log with the instructor; only completion is recorded.

The AI tutor will:

- Build curiosity in the lesson topic and motivate student interest.
- Ask diagnostic questions to assess anchor knowledge.
- Use that anchor knowledge to introduce new concepts, bringing each student toward a common baseline.
- Identify and correct misconceptions in real time through Socratic questioning.
- Ideally maintain an individualized knowledge record that tracks each student's interests, conceptual schema, and prior misconceptions across sessions.
- Provide structured feedback on the student's engagement at the end of the session, based on a provided rubric.

Worth noting as an instructor. The tutor is not only a delivery mechanism — it is an instance of the very interaction design the demonstration question investigates: Socratic questioning versus answer-giving as a determinant of learning versus offloading. Students arrive in class having just been a participant in the kind of study they are about to survey. Naming this recursion is worth one moment of class time: it makes the demonstration question personally consequential and gives students a concrete referent for abstract terms in the literature they will map.

D. Written Summary (~10 min)

After the video and the tutor session, students write a brief summary in their own words of the key terms and core concepts: the five steps and what each one delegates, and the difference between a search result and a generated summary. This is a submitted prelab item. Consolidating in writing is what makes the sequence available for *use* in class rather than merely recognized.

E. Tool Access Verification (~5 min)

Students confirm before class that their AI platform works and its deep research mode is available to them (including any usage limits); that ResearchRabbit opens; that their Google account opens NotebookLM; and that their library proxy login works.

F. Student Guide Review (~10 min)

Students skim the in-class steps in the Student Guide so they arrive with the sequence already in mind. The guide is deliberately not so detailed that a student could complete the investigation without the live class structure — the decisions students make in class (which question to commit to, which columns the matrix needs, which claim to spot-check) are the substance of the session.

Part 2 — In-Class Investigation

2 × 110 min

Students run the full workflow on their own research question while the instructor guides the class through each stage. The investigation proceeds in parallel as a class: students explore, hit friction, observe peers, iterate, and work toward a completed pass through the workflow with instructor coaching. The Student Guide provides the prompts, steps, and decisions. Students work in a single AI conversation across both class periods and keep it for later review.

Two logistical points the instructor coordinates. Students must start the deep research run early so it can work in the background — which may mean it sits between the two class meetings — and they should not open it until Stage 8 — seeing the automated answer early contaminates the judgment process we want them to experience in Steps 1-4.

A. Opening Activation for the Lesson (~15 min)

Discuss the importance of how new discoveries and innovation are built on the written literature that documents previous research and knowledge. Our access to this knowledge is essential to advancement. Describe the scale of the literature (roughly 2.8 million indexed articles in 2022, growing about 5.6% per year) and the historical arc of how we accessed it historically: memex 1945 → citation index 1964 → mediated online search → Google Scholar 2004 → AI-assisted synthesis. Introduce the great promise that AI offers in synthesizing massive amounts of information, and the perils of accepting that synthesis without verification.

B. AI Supported Literature Review: A Step-by-Step Workflow (~165 min)

AI assistance is integrated at each stage below, and the amount delegated grows as the block proceeds. Predictions and knowledge checks are solicited throughout, and peer discussion is central.

Stage	Move	What happens	Time
1	Refine the question	Students run the question-refinement prompt on the question they wrote by hand. The AI critiques and proposes sharper versions; the student chooses — their original, an AI suggestion, or a hybrid.	~15 min
2	Launch deep research	Students open their AI's deep research mode, paste the refined question, and start the run. It works in the background — which may span the break between the two class meetings — and is not opened until Stage 8.	~10 min
3	Step 1 — Claim survey	Students fill in the claim-survey prompt with their own question, hypothesis, primary field, and an adjacent field, then run it with web search on. The prompt explicitly stops short of synthesis, so the survey stays a list of papers and findings. Students skim the results, choose the two or three most relevant papers, and check full-text access. The class compares this automated process against the prelab Google Scholar experience: what sources did each process surface, and what output can be trusted for each.	~20 min
4	Step 2 — Citation mapping	Students seed ResearchRabbit with one central paper — from Stage 3 or from the prelab Google Scholar search — and explore the map: the cluster around the seed, foundational work, later work, and newer papers at the edges. The goal is awareness that this mode of exploring a literature exists, not mastery. A general-model prompt approximates it, returning up to twenty of the most relevant references, an optional graphical representation of the relationships, and a formatted PDF report.	~30 min
5	Step 3 — Synthesis matrix	Students build a five-source set (two prelab PDFs, two from Stage 3, one from Stage 4), decide which columns their question requires, and run the matrix prompt — which deliberately stops at the matrix and returns it as both an Excel spreadsheet and an interactive HTML view with expandable cells. Then the class pauses: five minutes, no AI, each student writes their own read of the matrix.	~30 min
6	Step 4 — Grounded synthesis and gap analysis	A two-part prompt: synthesize across the uploaded papers only, flagging blank cells that suggest gaps; then produce an independent 200-word assessment from web search, separating established, contested, and emerging. Students compare the two against each other and against the read they just wrote.	~30 min

Stage	Move	What happens	Time
7	Spot-check	Each student picks one specific, checkable AI claim about one paper, goes to the primary source, and records three lines: the claim as stated, what the source actually says, and the verdict (confirmed / overstated / wrong / could not access).	~10 min
8	Step 5 — Deep research	Students open the report launched in Stage 2 and review it with questions in hand: can they see the steps inside it, how do its references compare, what does it add? The step-comparison table itself is post-class work.	~20 min
The load-bearing five minutes. The write-down pause at the end of Stage 5 is where each student forms an independent read of their own evidence before the AI offers one. Run it as a timed, whole-class stop rather than leaving it to individual discretion, and make the prompt specific: where the papers agree, where they disagree, and which blank cells look like real gaps. Those are precisely the three judgments the Stage 6 prompt will produce.			

C. AI Tool for Exploration of Sources: NotebookLM (~30 min)

This segment sits outside the five-step workflow because it answers a different question. The five steps answer “what does the field know?” This asks “now how do I actually learn this material?” Mapping a field and understanding it are different activities, and the lesson separates them explicitly.

- **Instructor demo (~5 min).** The instructor opens a pre-built notebook on the demonstration question, asks one live question, and clicks one inline citation through to the exact source passage — making three points briskly: this is the payoff for verification, the corpus boundary is both the tool's virtue and its limit, and grounding reduces error without eliminating it.
- **Student exploration (~18 min).** Students build a notebook from the corpus they just curated — their matrix papers plus the deep research report if it has finished — and explore, clicking at least one inline citation to check the source passage. The asymmetry is named explicitly: the papers they have verified, the report they have not, so its claims are leads rather than established facts.
- **Feature share (~7 min).** Volunteers demo what they found, and the class closes the segment on two questions: what limitations did you hit in the tool, and what features would you add if you were designing one? Students are also encouraged to build notebooks for their other courses.

D. Reflect: Reading Your Field (~10 min)

Students close the block by writing one-line answers to three prompts; some are shared aloud.

- What is the consensus claim in your area — what does the field treat as basically settled?
- What is contested or unresolved — where are researchers still arguing, and how could you tell?
- What is a question you could investigate that no paper you found has directly answered?

The second prompt is the one to press if there is time: articulating *how you could tell* something was contested forces students to name the signals — multiple groups, conflicting results, hedged reviews, varying methods. Students bet on exactly this at the video's second pause and tested those bets in Stage 6; referencing that arc closes the loop. Students leave with a possible research gap, not a final answer, and the postlab is where they test it against an enhanced matrix. Not having a candidate yet is acceptable.

Deliverables

Before students leave, they save or export their AI conversation, their synthesis matrix, and their notebook link — the post-class work builds directly on all three.

Part 3 — Post-Class Work

~2.5 hours

Students read the deep research report closely if they did not finish it during class. They then complete a step-comparison table allowing them to consolidate their thinking about the workflow they conducted in class. The students then pull everything together in a literature research summary that ends with a gap statement and reflection on the process.

A. Review In-Class Work and Generate the Step-Comparison Table (~30 min)

Students read the deep research report closely if they did not finish it in class, then review their AI conversation from class with attention to the synthesis matrix and the Stage 6 synthesis. They complete a comparison table with one row for each move in the lesson — manual search, claim survey, citation map, synthesis matrix, grounded synthesis, spot-check, deep research report — and three columns: what it helped me find or do, what I verified, and what I would trust it for. The table is submitted as a PDF.

	What it helped me find/do	What I verified	What I would trust it for
Manual search (prelab)			
Claim survey (Step 1)			
Citation map (Step 2)			
Synthesis matrix (Step 3)			
Grounded synthesis (Step 4)			
Spot-check			

	What it helped me find/do	What I verified	What I would trust it for
Deep research report (Step 5)			

B. Literature Research Summary Document (~90 min)

This summary is the main deliverable for this lab. Students will revise their research question and then expand the research papers they identified and the synthesis matrix they created in class to include 6–10 sources. Five elements:

- **The question and its evolution.** Three stages — the prelab version written by hand, the AI-critiqued version from the start of class, and a final version after seeing the literature — with one line at each point on what changed and why.
- **Synthesis matrix, expanded.** Grown to 6–10 papers using citations already surfaced in earlier searches but not used in the first matrix. Full-text PDFs where possible; cells relied upon verified; unknown cells left blank.
- **A small map of the field.** 6–10 papers with one-sentence characterizations personally verified against the primary source rather than against an AI summary.
- **Gap statement.** Using the matrix to propose a research gap — the in-class candidate or a new one surfaced by the expanded matrix. Blank cells, disagreements between papers, and missing populations or conditions are the strongest support. A real gap should matter and should point toward evidence that would resolve it. If the matrix kills the candidate, that is itself a finding.
- **A reflection.** Which steps were most useful for which sub-tasks; what the notebook added that reading alone did not, including features classmates shared; whether the deep research report found the gap the student identified, missed it, or suggested something unconsidered; what surprised them; and how going to primary sources changed their understanding compared with relying on AI summaries.

C. Retrieval Practice / Course Review (~30 min)

Per the course's standard structure, students spend approximately thirty minutes on integrated course review using the course's customized AI-enhanced retrieval tool. This tool supports spaced repetition and interleaving of material across the full course, functioning as an adaptive alternative to traditional flashcards.

Assessment Approach

Consistent with the course's assessment philosophy, the process artifacts are graded, not just the product — the comparison table and the reflection are what reveal whether a student exercised judgment or transcribed AI output. Weighting is specified in the Instructor Guide: question quality 15%, synthesis matrix 20%, literature map 20%, gap statement 20%, step-comparison table 10%, reflection and verification 15%. Prelab and in-class process items are checked for completion rather than scored, and the tutor's engagement score carries no course grade.

Everything Students Turn In

Bold items are submitted; the others should be available to show the instructor on request. The prelab items are required before class because the in-class work depends on them.

Due before class, with the prelab

- The research question, in the version written without AI.
- Manual Google Scholar search notes, including the friction.
- The three-sentence deep-read summary and the one question that paper alone could not answer.
- Full-text PDFs of two or three central papers, or a note documenting the access routes attempted.
- Both written pause answers from the demonstration video.
- **The tutor session chat log, shared with the instructor. Only completion is recorded.**
- **The written summary of the five steps, what each delegates, and the difference between a search result and a generated summary.**

Due at the end of the in-class investigation

- The research question settled on at the start of class, with one line on what changed and why.
- The first synthesis matrix (five sources), with any abstract-only rows marked.
- The student's own written read of the matrix, written before seeing the AI's synthesis.
- The spot-check record: the claim as the AI stated it, what the source actually says, and the verdict.
- One-line answers to the three closing questions.

Due after class

- **The completed step-comparison table, as a PDF.**
- **The Literature Research Summary Document — all five elements, with the expanded 6–10 paper synthesis matrix.**
- **The NotebookLM notebook of curated sources, or a link to it.**

Companion Documents

Document	What it provides
Student Guide — Literature Research with AI	Student-facing preparation, investigation, reflection, and submission instructions. The authoritative source for timing, sequence, prompts, and deliverables; this overview and the Instructor Guide are both aligned to it.
Instructor Guide	Full pedagogical rationale, minute-by-minute in-class protocol, prompt templates (Appendix A), contested-finding signals (Appendix B), primary-source access routes (Appendix C), demonstration video specification (Appendix D), the evidence base for the design (Appendix E), and the assessment rubric.
Prelab Tutor Document	The eight-section tutor instruction set students upload to their AI for the Part 1 Socratic session, built on the course's AI-mediated prelab design guide.

Document	What it provides
Demonstration Video	Team-produced walkthrough of all five steps on the shared demonstration question, including the two written pause prompts. Link distributed with the prelab; specification in Instructor Guide Appendix D.
Instructor PowerPoint Deck	In-class support for pacing the stages and projecting the five-step table and prompts.