Original research article
UDC 811.112.2:92.09
DOI 10.21618/fil2633607a
COBISS.RS-ID 144490241

Stefano Adamo[1]
University of Banja Luka
Faculty of Philology
Department of Romance Studies

FROM SCRIBE TO ARCHITECT

A PRACTITIONER'S REFLECTION ON TEACHING ACADEMIC WRITING IN THE AGE OF LARGE LANGUAGE MODELS

Abstract: *This paper begins from a classroom observation: students in an academic writing course who began, mid-semester, producing literature reviews of unusual fluency — prose that was carefully structured, appropriately hedged, and, on closer inspection, partly fabricated. The observation points to a structural problem. For decades, instructors have treated writing fluency as a serviceable proxy for intellectual engagement; generative AI has broken that proxy, not by improving student thinking but by making polished academic prose available without it. The central claim of this paper is that AI amplifies existing competence rather than compensating for its absence: students who already know how to read critically, formulate an argument, and evaluate a source use AI productively; those who do not are helped to sound as though they had been. Against this background, the paper proposes a concrete three-phase workflow — built around structured prompting, distributed research across multiple AI tools, and progressive conversation in a purpose-built AI environment — designed to teach the literature review as an act of design rather than transcription. The larger implication is pedagogical: if the course can no longer evaluate writing as a finished product, it must evaluate the decisions that produce writing. Teaching academic writing, in the age of large language models, means teaching judgment.*

Key words: *academic writing; generative AI; large language models; literature review; pedagogy; AI literacy; authorship; higher education*

1. An Unexpected Elegance

I remember the moment with some precision. It was mid-semester, late in the spring term. I was grading a batch of papers submitted by students in my academic

[1] stefano.adamo@flf.unibl.org

writing course, fourth-year undergraduates, most of them still struggling with the twin challenges of synthesizing sources and imposing logical order on what they had found. The papers were typically what I had come to expect: earnest, uneven, often correct in their facts and confused in their arguments.

Several submissions that week were different. The prose was unusually fluent. Sentences were longer, more varied, more carefully balanced; transitions between paragraphs did their work smoothly; the vocabulary was precise in ways that surprised me. One student, who had turned in something barely legible six weeks earlier, now wrote with a confidence that made me pause and reread the first page.

My first reaction was pleasure; the particular pleasure of a teacher who believes the thing has finally worked. Then came the doubt. It arrived, as such doubts tend to, as a small and nagging sensation rather than a clear accusation: *was this her work?*

I want to be careful about how I frame that question. The students had not necessarily done anything wrong; institutional rules on AI use were still being debated; several of them had probably simply followed the path of least resistance in a world where that path now runs through a large language model. The question was not, in the first instance, a disciplinary one. It was a pedagogical one; perhaps even an epistemological one. If a student submits polished prose and I cannot determine from the prose itself whether it was produced by the student or by a machine, then what, exactly, have I learned about the student from reading it?

That is the question this paper tries to answer, not definitively, but practically. It begins from this classroom moment not because classroom observations are probative evidence, but because they are where teachers live, and because the unease I felt at that grading table is, I suspect, widely shared. Something has changed in the relationship between student and text. Understanding what has changed, and why it matters, requires attending carefully to what writing actually does, and what it stops doing when it is produced without the effort that normally produces it.

2. The Problem of Fluency Without Understanding

Academic writing instruction has always rested on an assumption: that the quality of student prose is a serviceable proxy for the quality of student thought. It was never a perfect proxy — a talented stylist can produce elegant sentences without clear ideas, as any reader of certain literary theory will attest — but it was a reasonable one. A student who could build a well-structured paragraph, manage subordination without losing the thread of the argument, and deploy technical vocabulary with precision had almost certainly been doing some thinking, whether

or not the thinking was correct. Fluency was not the same as understanding, but it was plausible evidence of engagement with difficulty.

What generative AI has changed is not the availability of fluent prose. It has changed its cost. Producing a coherent, well-organized, appropriately hedged paragraph in academic register now requires approximately thirty seconds and a competent prompt. The intellectual labor that used to produce fluent writing, that is to say reading carefully, selecting what matters, formulating a claim, testing it against an objection, and so forth, can now be bypassed entirely. And this changes things profoundly, because that labor was not merely the means to an end. It was, in large part, the education.

Nowhere in my course was this transformation more visible than in the literature review. Of all the writing tasks I assign, the literature review has always been the hardest to teach. Students are asked to do something genuinely difficult: not merely to locate sources, but to read them carefully enough to understand where they agree and where they diverge, and to build an argument about the current state of a conversation in the field. Most students, before the widespread availability of AI tools, managed something considerably more modest. With few exceptions, the best of them produced a loosely organized sequence of summaries tethered to a list of weblinks rather than to properly cited scholarly sources, with no real sense that the texts they were describing were in any kind of dialogue with one another. A literature review, in the full academic sense of the term, was beyond the reach of most of them.

Then I began receiving submissions that looked entirely different. The following passage, drawn from a student assignment submitted mid-semester, is representative of what changed:

> De Rogatis evidenzia come l'iniziazione di Delia passi attraverso un confronto con il corpo materno e i suoi segreti, in una dinamica che richiama riti antichi e miti classici (De Rogatis, 2024: 47). Nel romanzo di Di Pietrantonio, la malattia della madre genera un dialogo silenzioso ma emotivamente intenso: "Mi parlava coi gesti, con le mani. La voce, quando usciva, era debole, sempre lenta" (Di Pietrantonio, 2023:p.45). Il silenzio diventa una presenza eloquente, e la narrazione funziona come strumento catartico per attraversare il dolore e ridefinire l'identità. Come afferma Sambuco, la scrittura femminile mira a dare senso al dolore, non solo elaborando il lutto, ma riordinando ciò che è stato spezzato attraverso evocazioni sensoriali e ricordi corporei (Sambuco, 2012:p.56). Il ritorno nei luoghi dell'infanzia — Napoli per Delia, l'Abruzzo per Amanda de *L'età fragile* — non è solo geografico ma simbolico, un movimento verso un sé

> più consapevole. Riccardi sostiene che l'io narrante femminile rappresenta un dispositivo di autodeterminazione, capace di riscrivere il trauma e far emergere nuove soggettività (Riccardi, 2018:p.81). La narrazione, quindi, partecipa attivamente alla ridefinizione dell'esperienza. Le protagoniste si muovono in una tensione costante tra appartenenza e distacco, tra eredità materna e bisogno di emancipazione, trasformando la parola in forma di resistenza e affermazione.[2]

The passage is, on first reading, impressive. The student has done something that most undergraduates cannot manage without considerable coaching: she has integrated secondary sources into her own analytical prose rather than merely listing or summarizing them. The citations arrive as natural extensions of an argument rather than decorative appendages to it. The orchestration of multiple scholarly voices — De Rogatis on myth and initiation, Sambuco on feminine writing as a response to grief, Riccardi on the female narrative self as a form of self-determination — creates the impression of a student who has read widely, thought carefully, and knows how to move between a theoretical framework and a close reading of specific texts. It is, in short, what a literature review is supposed to look like.

A closer reading reveals something else. None of the verifiable anchors in the passage withstand scrutiny. The three citations to secondary scholarship each fail in a different way. De Rogatis (2024) is a real article; it does appear in the journal *Allegoria*. But it appears in issues 69–70, beginning on page 273. The page number cited (47) belongs to a different article entirely. The student is citing a page she has never read, in an article whose argument she cannot have reconstructed from the passage attributed to it. The problem is not only that the citation is wrong; it is that the wrongness is invisible to any reader who does not happen to have the volume

[2] "De Rogatis highlights how Delia's initiation unfolds through a confrontation with the maternal body and its secrets, in a dynamic that evokes ancient rites and classical myths (De Rogatis, 2024:p.47). In Di Pietrantonio's novel, the mother's illness generates a silent but emotionally intense dialogue: 'She spoke to me with gestures, with her hands. Her voice, when it came, was weak, always slow' (Di Pietrantonio, 2023:p.45). Silence becomes an eloquent presence, and the narrative functions as a cathartic instrument for processing pain and redefining identity. As Sambuco argues, feminine writing aims to make sense of pain, not only by working through grief, but by reordering what has been broken through sensory evocations and bodily memories (Sambuco, 2012:p.56). The return to childhood places — Naples for Delia, Abruzzo for Amanda in *L'età fragile* — is not merely geographical but also symbolic: a movement toward a more conscious sense of self. Riccardi maintains that the female narrative self represents a device of self-determination, capable of rewriting trauma and allowing new subjectivities to emerge (Riccardi, 2018:p.81). Narration, therefore, actively participates in the redefinition of experience. The protagonists move in a constant tension between belonging and detachment, between maternal inheritance and the need for emancipation, transforming language into a form of resistance and affirmation." (Author's translation of an unedited term paper submitted in the academic year 2024/2025).

at hand. Sambuco (2012) presents a different set of difficulties: the book exists, but the student's bibliography gives Florence and *Le Lettere* as place and publisher, when the book was in fact published in Padua by *Il Poligrafo*. Page 56 of the actual volume contains nothing resembling the claim attributed to it. Riccardi (2018) is the most instructive case of all. The student's bibliography presents it as a journal article; it is not. The cited review, *Studi italiani*, is a tellingly vague reference: more than one journal bears that name. It is instead a doctoral dissertation defended in 2017 at the Ludwig Maximilian Universität in Munich, where it remains available on Academia.edu. Page 81 of the dissertation, the page on which the student informs us that Riccardi develops her argument about the female narrative self as a device of self-determination, contains the candidate's curriculum vitae. Finally, even the direct quotation from Di Pietrantonio's novel is not authentic: the passage does not appear on page 45 of the edition cited, nor anywhere else.

What the student has produced, in other words, is the architecture of a literature review: a confident scholarly register, correctly formatted in-text citations, arguments that appear to be grounded in specific passages from named sources. The architecture looks convincing. The foundations, upon inspection, are not there.

What the passage illustrates in concrete terms, the cognitive scientist Steven Pinker's account of language and thought explains in theoretical ones. Composing a sentence, Pinker has argued with characteristic precision, forces a writer to commit to a sequence, namely, to decide what is cause and what is effect, what is claim and what is evidence, what the reader already knows and what they need to be told (Pinker, 2014, chap. 4). These commitments are not decorative; they are constitutive of understanding. You do not write the sentence and then understand the argument. You understand the argument by writing the sentence. In a sense, the process is the product.

Large language models do not work this way. They are systems of statistical prediction: they generate text by calculating probable continuations of a given sequence of tokens. They produce sentences that are grammatically well-formed and stylistically recognizable because such sentences appear frequently in their training data, not because they represent any speaker's attempt to model reality or reason through a problem. The prose that results can be remarkably fluent. It can also be, as it often is, generically competent, strategically hedged, epistemically hollow. It reads like thinking, without doing any.

From a pedagogical standpoint, this produces an uncomfortable paradox. If we evaluate student writing as a product, that is, if we assess the sentences, the paragraphs, the structure, the vocabulary, and so forth, then AI-assisted writing can

pass nearly any test we currently apply. The text looks the way a literature review is supposed to look: it uses the right verbs (*argues*, *highlights*, *maintains*), uses the right register, and avoids the most obvious failures of logic and grammar. What it cannot replicate is the sediment of understanding that difficult writing leaves behind in the person who struggled to produce it. The student who has written a sentence, recognized it as wrong, deleted it, and tried again has acquired something that no volume of polished AI output can substitute: a felt sense of what the argument requires, and what it will not support.

My classroom observations, impressionistic as they necessarily are, suggest that the gap the passage above makes visible on the page is equally visible in spoken exchange. Students who relied on AI to generate their literature reviews were consistently less able, in subsequent seminars, to defend the positions that prose appeared to take. They knew what the text said; they did not know why it might be true, or why an informed reader might push back. The writing had been produced without the thinking the writing was supposed to represent. The thinking, accordingly, had not taken place.

This is not, it bears emphasizing, an argument against AI in academic settings. It is an argument about what AI does and does not do, and about the pedagogical cost of mistaking its outputs for evidence of student understanding. A fuller account of that cost and of how it might be addressed requires examining what the research literature has established about the way students actually use these tools, and what happens when they do.

3. What the Literature Tells Us: A Selective Review

Three years of intensive scholarly attention have produced a literature on generative AI and academic writing that is, by now, both substantial and usefully convergent. Researchers and editors have approached the question from different disciplinary angles, such as publishing ethics, empirical education research, cognitive science, information literacy, and reached positions that are neither panicked nor naively optimistic. What follows is a selective reading organized around three themes. Its aim is not comprehensiveness but analytical precision: to establish what the evidence actually says about three problems that bear directly on the pedagogical argument of this paper.

AI as tool, not author

The most immediate question the emergence of generative AI posed for scholarly publishing was institutional before it was pedagogical: could an AI be listed as an author of a paper? The answer, across an unusually wide range of journals and publishers, has been consistently negative, and the reasoning behind that consensus is more philosophically interesting than it might first appear.

The objection turns on a specific and defensible understanding of what authorship means in academic contexts. Scholarly writing is an accountable act of judgment: the author stakes a claim, takes responsibility for the evidence adduced in its support, and can be held to account for everything they have said. An AI system cannot be held accountable in any meaningful sense. It cannot revise its position in response to sustained criticism, acknowledge that it was wrong, or bear moral responsibility for the consequences of a mistaken assertion. Kaebnick and colleagues argued in 2023 that this asymmetry of accountability is not a technical limitation that further development will eventually overcome; it is structural to what these systems are. They can produce the text of a scholarly argument. They cannot bear the responsibility that scholarly argument requires (Kaebnick *et al.*, 2023).

This position has achieved a degree of consensus rarely seen in such a rapidly evolving field. A systematic analysis by Perkins and Roe (2024) of editorial policies across twenty-eight publishers found consistent agreement on the core principle: AI tools may legitimately assist in the preparation of academic work, but they may not be named as authors. Hosseini and colleagues (2023), examining disclosure norms across research contexts, pointed toward the corresponding corollary: any substantive use of AI assistance must be declared, so that readers can evaluate the epistemic status of what they are reading. Leung and colleagues (2023) reached similar conclusions from the standpoint of research integrity and editorial practice. The consensus is not a prohibition on AI use; it is a clarification of where intellectual responsibility lies. The author remains the author, and they must be identifiable and morally implicated in the claims they submit for scrutiny.

Who benefits, and how

The empirical literature has moved beyond the question of whether AI should be used to the more tractable question of how it is actually used, and with what effects. The findings here are more differentiated, and considerably more instructive for the teacher in a writing classroom.

Nguyen and colleagues (2024), in a study published in *Studies in Higher Education*, documented a pattern that has since been extended by subsequent research. At the doctoral level, students with stronger baseline writing competence tend to interact with AI generatively and iteratively, using it to test formulations, probe for counterarguments, and refine positions they are already capable of articulating. Students with weaker foundations tend to use AI passively: accepting outputs without critical engagement, inserting them with minimal revision, and performing no better, and often worse, than they would have without any AI assistance at all. The difference is not one of access or willingness. It is one of competence. To interact critically with an AI output, you must already know enough to evaluate it. You must be able to recognize when a generated paragraph has missed the point, softened a claim that should be sharp, or produced a plausible-sounding sentence that is, on inspection, empty.

Kim and colleagues extended this line of inquiry across two studies (2025, 2026), examining the relationship between what they term AI literacy — the capacity to prompt effectively, evaluate outputs critically, and integrate AI assistance into a coherent intellectual process — and the quality of the resulting academic writing. Their findings point consistently in the same direction: AI literacy is not a standalone skill that can be layered onto a student who lacks foundational competence in academic argument. It presupposes that competence. Lin's (2024) model of human-AI collaboration arrives at a parallel conclusion: effective collaboration requires a form of metacognitive awareness, that is the ability to monitor one's own reasoning and identify where a tool is genuinely helping and where it is substituting for thought that should be the writer's own. That awareness is itself the product of prior intellectual formation. It cannot be acquired by using AI; I would argue that it is what makes AI use productive rather than merely comfortable.

The risks

The third theme in the literature concerns what goes wrong, and the evidence here is specific enough to require careful attention. The popular conversation about AI in education tends to foreground plagiarism, the fear that students will pass off AI-generated text as their own. This is a real concern, but it is also, in a certain sense, the wrong one to lead with. The more serious risk is epistemic rather than disciplinary.

AI language models hallucinate: they generate plausible, fluent, grammatically impeccable text that is factually incorrect, and they do so without any indication

that anything has gone wrong. In academic writing, this takes a particularly consequential form. Cheng and colleagues (2025) documented high rates of fabricated citations in AI-generated research outputs, papers that do not exist, authors who did not write them, journals that never published the works attributed to them. What makes this so insidious is precisely the fluency that makes AI prose so convincing: a student who cannot independently verify a source list will not notice that part of it is fictional. Pinker's warning, developed in the previous section, that statistical competence does not constitute semantic understanding now has concrete empirical backing. A model that generates authoritative-sounding citations is doing what it always does: computing probable continuations. It has no idea whether the paper it has just named exists.

Two further risks are harder to quantify but deserve acknowledgment. The first is stylistic homogenization: academic prose generated by AI tends toward a recognizable generic register. It is usually correct, hedged, logically organized, and tonally undifferentiated. As more student work passes through the same underlying models, the range of voices in academic writing narrows, and with it the range of intellectual personalities that academic writing has historically cultivated. The second is skill atrophy: the concern, grounded in general principles of cognitive science if not yet in longitudinal data specific to AI use, that capacities not exercised will weaken. A student who never has to struggle with formulation is likely to become less capable of formulating.

What the literature does not provide

The research has converged, then, on a sober and differentiated position. AI must not be treated as an author; it should be used transparently and with declared limitations; its practical benefits accrue mainly to those who already possess the competence to use it well; and its risks, whether epistemic, stylistic, or developmental, are real and documented. This is a more honest and more useful consensus than either the panicked prohibition or the uncritical enthusiasm one sometimes encounters in public debate.

What the literature does not yet provide is a pedagogically actionable model: a concrete account of how an instructor might actually redesign a writing course to translate these principles into practice. The research diagnoses the problem; it offers relatively little guidance on what to do about it on a Tuesday morning in a room full of undergraduates. It is that gap this paper attempts to address.

4. Authorship Reconsidered: The Architect and the Mason

The Italian passage I examined in the previous section illustrates a problem that editing can expose but cannot fix. Its author had produced a literature review she could not defend: citations she had not read, arguments whose sources she could not reconstruct, a scholarly voice that was nowhere hers. The difficulty was not carelessness; the text is, on its surface, carefully made. The difficulty was that the author of the text, in any robust sense, was not the student. Someone — or something — else had done the writing. The student, at best, had placed the order.

The philosopher Luciano Floridi has given us a vocabulary for describing what this new division of labor looks like, and for asking whether it can be made to work. In his essay "Distant Writing" (2025), Floridi argues that AI-assisted composition reorganizes the relationship between human and text along lines that are old in architecture but new in prose. The author of an AI-assisted document is no longer the sentence-producer. She becomes a *meta-author*: a designer who sets constraints, specifies ends, evaluates candidate outputs, and accepts or declines them. The model performs the execution. The author is the architect; the model is the mason. The building is the architect's; the labor of laying the stones is not. Floridi has returned to the same architect/mason opposition, in more explicit terms, in a recent podcast interview (The BSMT by Gianluca Gazzoli, 2025, 41:39-42:17).

This is a tidier reconfiguration than it first appears. Floridi's division of labor is not a separation of prestige from drudgery; it is a redistribution of what the agent is and is not responsible for. An architect who cannot read an engineer's drawings — who cannot tell, by inspection, whether a proposed beam will carry the load his design places on it — is not an architect. He is a client. The metaphor is only consoling for those who bring real competence to the authorial role. Without that competence, "meta-authorship" is abdication dressed in more dignified language.

Floridi develops the philosophical ground for this worry in an earlier piece, significantly titled "AI as Agency Without Intelligence" (2023). Large language models, he argues there, *act* effectively, i.e., they produce text that achieves its communicative function in most contexts, without anything that could be called understanding. They have agency, in the narrow sense that their outputs have consequences; but they lack any semantic grasp of what they produce. What we are dealing with, then, is something genuinely new in the writer's toolkit: neither a tool in the neutral sense — a pen does not shape the meaning that emerges from its ink — nor a collaborator in any meaningful one, since collaboration presupposes that both parties understand the work. It is, rather, a high-capacity executor: fluent, yes,

but fundamentally unable to know whether what it has written is true. Its outputs cannot be trusted the way a competent human's can, because trust, in scholarly writing, rests on understanding, and understanding is precisely what the model does not have. The author who delegates to such a system therefore remains the only point at which the text's meaning is actually apprehended, if it is apprehended at all. From this single fact, almost every serious warning in the literature reviewed earlier follows directly: hallucination, because the model cannot check its claims against reality; stylistic flattening, because it optimizes for statistical plausibility rather than intellectual voice; and, most consequentially, the displacement of thinking by the performance of thinking, because fluent prose no longer requires or implies that someone has thought.

The pedagogical consequence Floridi does not draw, and which I want to press, is this: to be a good architect you must already understand engineering. The architect's role is not to compute the load on the beam; it is to know that the load matters, to recognize when a calculation looks suspicious, and to refuse a solution that does not meet the specification. That prior competence is what makes the delegation safe; not merely efficient, but intellectually honest. Remove it, and what looks like authorship is something else: a confident acceptance of outputs that the putative author is not in a position to evaluate. The student who submitted the fabricated citations I examined earlier had performed this role exactly. She had accepted a mason's work without possessing the engineering knowledge that would have let her see that the building could not stand.

An objection presents itself here, and it deserves a direct answer. Does the patient who consults a doctor not also delegate to an expert she does not understand? The analogy is tempting but misleading. The patient is responsible only for choosing to seek treatment; the doctor is responsible for the treatment itself, and is accountable for what it contains. An author, by contrast, is accountable for the content of the claim, not merely for the decision to submit it. Medical delegation distributes responsibility between two accountable agents; authorial delegation cannot, because the AI that drafted the text cannot be held to account for what the text asserts. What issues from the model, the author must own; and to own it responsibly, the author must be in a position to judge it. Engineering knowledge is not optional for the architect.

If that is right, then a further consequence follows, and it runs against one of the most popular arguments for AI in education. Generative models, it is frequently claimed, will *democratize* scholarly production: students who have struggled with the mechanics of academic prose will now produce work of genuine quality, and the old barriers of class, language, and educational background will at last begin to

fall (Heaven, 2023; Villasenor, 2025). There is something in this hope. Students whose first language is not English can now draft with a fluency that once would have required years of stylistic training. But the hope rests on a confusion. What AI democratizes is the *appearance* of competence, not competence itself. A student who cannot yet tell a sound argument from a plausible-sounding one is now able to produce text that a hurried reader, or the student himself, will mistake for one. The evidence reviewed earlier corroborates the point: the students who benefit most from AI are the already-competent ones. The others are helped to sound as though they had been helped.

The implication for pedagogy is not that AI should be prohibited. It is that the foundations of academic writing — reading carefully, distinguishing claim from evidence, recognizing when an argument has been made and when it has merely been asserted — become *more* urgent to teach, not less. If the meta-author is to be anything more than a rubber stamp, she must be formed first as something more modest: an author who has herself laid a few stones and felt their weight. What such a formation might look like in practice is the subject of the next section.

5. A Practical Proposal: Teaching Writing as Design

5.1. Phase 1: Building and Modifying the Research Prompt

A word on the example before I begin. The workflow I describe presupposes prior acquaintance with the subject. It is not a primer, and it is not a substitute for the reading, viewing, or fieldwork through which one first becomes familiar with a body of material. A student who had never watched a mafia film or read a page of film criticism could run the sequence I am about to describe, but the outputs would have nothing to bite against. She would be unable to tell which sources are canonical and which marginal, which gaps are real and which are artifacts of her own unfamiliarity with the field. The workflow works on prior material; it does not supply it. This is, in another register, the point of the previous section: AI is a tool for those who already know, in outline, what they want.

I will use, as a running example throughout the section, a literature review on the representation of the mafia in Italian cinema from 1998 to the present. I use this example because students working in a field of this kind typically arrive with some prior purchase: most have seen a few of the films, recognize some of the names, have formed some initial impressions. The workflow, in their hands, has something to act upon.

With that caveat stated, the sequence begins. I ask the student to articulate her research question as vaguely as she likes. The vagueness is the starting point, not the problem. If she cannot yet say what she wants to find out, she is at the beginning of a real research process; if she can say it precisely, she is further along than most undergraduates.

I then ask her to transform that vague question into a structured prompt using a framework called ACTIF[3] — a five-part checklist, glossed here in English, covering five elements: *Action* (what the AI is being asked to do), *Context* (where the request sits, what project it serves), *Tone* (the register of the output), *Identity* (the persona or expertise the AI is being asked to assume), *Format* (how the output should be structured). The acronym is functional: it gives students a checklist that covers the variables they most often leave implicit when they prompt, and which the AI, left to its own defaults, will fill in according to statistical plausibility rather than scholarly need.

Here is how the framework operates in practice. A student beginning a literature review on the representation of the mafia in Italian cinema starts with a formulation such as: *I want to know how the representation of the mafia has changed in Italian cinema.* An initial ACTIF prompt restates this along the five axes:

- **Action:** Generate an annotated critical overview of the secondary scholarship on the representation of the mafia in Italian cinema from 1998 to the present, organized by themes and identifying the main argument each source advances.
- **Context:** Preparatory bibliography for a seminar paper in Italian film studies.
- **Tone:** Academic, analytical, comparative.
- **Identity:** Film scholar specializing in contemporary Italian cinema and its relationship to Italian cultural and political history.
- **Format:** Structured overview with thematic sections and full bibliographic references.

This is a recognizable prompt, and it would produce a recognizable response. But it is not yet the prompt I want the student to use. The next step, which is where the pedagogical work is done, is to modify the prompt so that it asks the AI to do *only* what the AI is actually good at, and reserves for the student the cognitive operations that must remain hers.

[3] ACTIF is an acronym drawn from French-language prompt-engineering practice. A fuller description of the framework is available at https://prompting.numedu.org/.

The modification strips interpretation and synthesis from the brief and keeps only factual retrieval. Where the original asks for a *critical overview*, the modified version asks for an *annotated inventory*; where the original asks for *themes*, it asks for *topics named in peer-reviewed scholarship*; where the original asks for *argument*, it asks for *the scholarly position each source advances*. This way, the student must decide, line by line, which operations belong to the AI's competence and which to her own. She must notice that *significance* is interpretive and *publication date* is not; that *contested* is a judgment and *cited by* is a fact. A student who cannot make these distinctions is not yet ready to use AI in a research process. A student who can is already performing the analytical work the literature review will require.

The resulting modified prompt might read:

> 'Provide an annotated overview of Italian-language films on the mafia from 1998 to the present, identifying directors, titles, release dates, and the peer-reviewed scholarly works that discuss them. For each source, identify the author, the full title, the publication venue, the year, and the specific scholarly position advanced in the author's own terms, where possible. Preserve active hyperlinks to original sources.'

I teach this step because it is the only point in the workflow where the student is asked to discriminate, in an explicit way, between what she is delegating and what she is keeping. If she skips it, the rest of the workflow has no cognitive anchor; she will have built a research process on a prompt whose distribution of labor she has not examined.

5.2. Phase 2: Distributed Research

Once the modified prompt has been written, the second phase begins. The temptation at this point is to paste the prompt into whichever AI tool the student happens to have open, accept the output, and proceed. The workflow refuses this shortcut. It asks the student to run the same prompt across three distinct research tools and to hold the three outputs side by side. I currently ask my students to use Consensus, Perplexity Academic, and Gemini Deep Research,[4] but the specific tools matter less than the principle they enforce. Each draws on a partially different index of the scholarly literature; each composes its results through a different reasoning chain; each, in consequence, produces a different sample of what it takes

[4] The specific tools change quickly, and by the time this paper reaches print some will likely have been renamed, merged, or superseded. The triangulation principle is tool-agnostic: any three research assistants that index different corpora and reason in different ways will do the same pedagogical work.

the relevant scholarship to be. The three outputs will agree on some sources, diverge on others, and, most usefully, disagree about what is central and what is peripheral. That disagreement is the phase's epistemic payload.

Triangulation, as I use the term here, is not a mechanical cross-check. It is the habit of treating any single AI output as a *sample* rather than a *survey*, one reading of the field rather than the field itself. When Perplexity Academic names a monograph that neither of the others mentions, the student must decide whether it is a genuine discovery or an artifact of that tool's particular weighting. When two tools converge on a source and the third does not, the convergence is evidence of something real in the literature; the absence is evidence of something about the third tool. Students who run this phase for the first time almost always arrive surprised at how little the three reports overlap. That surprise is pedagogically useful: it unseats the assumption, quietly held by most undergraduates, that *the AI* (in the singular) has told them what the scholarship says.

The practical instruction at this stage is mundane: export each of the three reports with hyperlinks preserved. A literature review whose citations are not linked back to their sources has no epistemic spine: the student cannot verify a quotation, cannot check a page number, cannot confirm that a paper attributed to a given author actually exists. The fabricated citations I examined earlier in this paper were possible in part because the student had accepted an AI output in which the link between claim and source had been silently dissolved. Preserving the links forces the source back into view. It is the phase's small, concrete guard against exactly the failure mode that has made AI-assisted writing so epistemically fragile.

5.3. Phase 3: Structured Conversation in a Claude Project

The three exported reports are the raw material for what follows. They are not yet a literature review — they are three overlapping samples of one — and the move from sample to review is the cognitive work the student still has to do. For that work I ask the student to create a dedicated Claude Project. The choice of platform is not essential; any conversational AI that supports a stable custom system prompt and attached documents would serve. What matters is that the student works inside a persistent environment that has been configured, in advance, for the specific scholarly domain at hand.

The first thing the student writes inside the new Project is its system prompt. Writing the system prompt is itself an act of prior authorship: the student must specify the domain, the register, the kinds of claims that count as evidence, and

the standards by which a proposed interpretation should be judged. A student who cannot write this prompt cannot use the Project productively, because she will not yet know what she is asking the assistant to be. For the running example, the prompt might read:

> 'You are a research assistant specializing in contemporary Italian mafia cinema, supporting a scholar working on a literature review covering the period from 1998 to the present. You are familiar with the aesthetic, sociological, and historical literature on this corpus. Your role is to help analyze sources, identify patterns, suggest interpretations, and assist in drafting academic prose. Maintain a scholarly register throughout. Flag interpretive uncertainties explicitly. Distinguish between well-established critical positions and emerging or contested ones. Do not invent citations; if a claim cannot be anchored in the attached materials, say so.'

The final instruction is a direct response to the failure mode documented earlier in this paper: the model's willingness, absent an explicit prohibition, to invent plausible sources.

With the three exported reports attached to the Project, the student runs a sequence of three prompts, in strict order.[5] The first, *Mapping*, asks the assistant to produce a systematic map of the collected literature — thematic strands, disciplinary provenance, historical periods, film corpora, peer-reviewed versus non-peer-reviewed sources. The register is descriptive and organizational: the assistant is laying out what is there. The second, *Tensions*, is run only once the map has been validated; it asks the assistant to identify the main fault lines across the mapped literature — disagreements between aesthetic and sociological readings, between historical and contemporary approaches, between Italian and non-Italian critical traditions. The register is analytical and comparative: the assistant is surfacing divergence. The third, *Gaps and Questions*, is run only once both the map and the tensions have been validated. It asks for robust points of convergence, methodological or thematic gaps, and — in closing — three original research questions

[5] An alert reader will note a tension with the argument of §4. There I argued that the AI cannot perform interpretation, and in Phase 1 I built that conclusion into the prompt-modification step by stripping interpretive operations from the brief. The sequence described here appears to ask the AI to do exactly what I said it could not: identify tensions, detect gaps, generate questions. The appearance is misleading. What the AI does in Phase 3 is compose candidate interpretations from a corpus the student has already curated. It proposes; the student disposes. Every output is read, evaluated, and either accepted, revised, or rejected before it is allowed to shape the next step. The AI is no longer operating as an oracle — it is operating as a drafting assistant that accelerates the student's cognitive work without performing it on her behalf.

that the current literature has not adequately addressed. The register is evaluative and generative.

Each step's output is the next step's input, and each requires the student's validation before the sequence continues. A student who skips directly to Gaps and Questions will get a plausible-looking answer, but the answer will float free of the literature it purports to be about; the AI will have generated research questions from its own generic priors rather than from the specific corpus the student has assembled. The sequence works when the student is forced, at each stage, to pause and check whether what the assistant has produced matches what she has actually read.

For students who benefit from non-linear navigation of sources, I sometimes suggest, as a supplement rather than a substitute, uploading the exported reports and a selection of full-text PDFs into NotebookLM. This tool permits quick interrogation of a fixed corpus without generating open-ended extrapolations. It is useful for verification and for targeted rereading. I mention it here for completeness. The core workflow does not require it. What the workflow does require is a clear answer to the question of what, exactly, it is teaching. That is the subject to which I now turn.

5.4. What the Workflow Does, and Does Not, Teach

It would be a mistake to describe the workflow I have just outlined as a method for teaching academic writing. It is, more modestly, a method for teaching the *preparation* that academic writing in a field of secondary scholarship requires: the clarification of a research question, the assembly of a reading list, the identification of fault lines and gaps. What it teaches, at its most ambitious, is the habit of precision: the habit of asking, at each step, what exactly is being claimed, on what evidence, and with what confidence. That habit is not writing. But without it, writing has little to be about.

Nor should the workflow be mistaken for architecture. It is scaffolding: provisional, removable, erected to make a particular kind of work possible, not to replace it. The scaffolding carries the student as far as the threshold of the writing moment and then, properly, comes down. It does not teach her how to formulate a claim that is hers rather than the literature's. It does not teach her how to choose, from among the research questions the sequence has generated, the one that is worth spending a paper on. It does not teach her how to write a sentence she has not yet heard anywhere else, whose rhythm is hers, whose commitments she is prepared to defend. These are the operations that the workflow has carefully preserved for her,

because they are what writing actually is. What the workflow deliberately leaves to the student, such as the argument, the voice, the responsibility, is the subject of the section that follows.

6. What Remains Irreducibly Human

The argument, the voice, the responsibility: these are what the workflow has carefully preserved for the student to bring. Consider, first, the argument. At its most successful, the workflow ends with three research questions the student did not have at the beginning. But the questions are candidates, not conclusions. Selecting among them requires a judgment that no prompt can specify: which of these is worth pursuing, in this field, at this moment, given what this particular student has actually read, thought about, and can defend? The AI generates plausible questions. It cannot know which question is yours, or whether you have anything real to say about it. That knowledge, and that honesty about its absence, is what the choice of focus requires.

Voice is more elusive, but the point is the same. Every sentence in an academic argument carries commitments: about what counts as evidence, about how strong a claim the writer is prepared to stand behind, and about how much uncertainty she is willing to leave visible. These commitments are the writer's own, and they are audible; perhaps not always in the prose on the page, but reliably in the seminar room or in an oral exam. A student who has genuinely thought about the films she has been watching will write differently from a student who has assembled a list of sources she has not read. The difference may not show in fluency, but it is unmistakable the moment she is asked to say, in her own words, why any of it matters.

The deepest of these responsibilities connects most directly to the fabricated citations examined at the outset: the responsibility toward sources. In the scholarly tradition, a citation is a promise — a promise to every future reader that a real person, writing in a real venue, said something that can be verified. Every citation a student carries over from the workflow's output becomes a promise made under her own name. That means checking every linked source, reading every retained claim, and removing anything she cannot verify. These steps constitute the basic condition of intellectual honesty that authorship has always required. The AI, incapable of bearing responsibility for what it asserts, cannot fulfil this responsibility in her place.

What this implies for the classroom is a shift in what the course is producing. Not competent texts, but competent judges: students who can discriminate, evaluate, and decide what to stand behind. Assessment must follow this shift. An oral

defense of a literature review reveals in five minutes what a polished written submission conceals: whether the student knows why the argument holds, or merely knows what the argument says. A process portfolio — prompts, AI outputs, annotations, and revision history included — makes the intellectual decisions visible rather than hiding them behind a finished surface. In each case, what is being evaluated is no longer the finished text but the capacity that stands behind the argument.

7. Conclusion: Back to the Classroom

I return, at the close, to the question with which this paper began. The student whose prose had suddenly become fluent: was this her work? The honest answer, now, is: some of it was. The formulations were not hers; the sentence rhythms were borrowed from a model trained on a million texts; the citations, at least some of them, were invented. But if she had spent time with the literature, done the reading, formed some initial impressions, and then used those materials to shape what she asked the AI to produce, then the questions she brought to the workflow, the sources she chose to keep, the judgment about what was worth pursuing: those were hers. The boundary is neither clean nor stable. It moves depending on what she brought to the conversation with the machine.

The more important question, though, is not whether the work was hers. It is whether she learned anything by producing it. And here the answer is harder. If the course was not redesigned around the decisions that produced the text (no oral defense, no process portfolio, no one asking her, at each stage, to justify the choices she was making) then she may have learned almost nothing. She has a polished text. She does not necessarily have what the text was supposed to leave behind: a felt sense of the argument, a familiarity with the sources, an ability to defend a position under challenge.

What has changed in my own practice is not the final product I ask for. It is what I watch while the product is being made. The course still ends with a term paper that includes a literature review, however brief. But the paper is no longer the destination — it is the trace left by a series of decisions I am now in a position to assess. How did she formulate the research question? How did she modify the prompt? Which sources did she verify, which did she discard, and on what grounds? The writing is evidence, but evidence of something the writing itself no longer contains.

This is, in some respects, a return to something older. For centuries, the rhetorical tradition taught composition as a discipline of thought: a way of clarifying what you believed, testing it against objections, and presenting it in a form an audience

could engage with. The AI has not destroyed that tradition. It has stripped away everything that surrounded it, the fluency, the formatting, the register, and left the core exposed, harder to fake and harder to grade.

Academic literacy in the next decade will increasingly mean the ability to design intelligent conversations: to know what to ask, how to evaluate an answer, and when to trust your own judgment over a fluent machine. That is a more demanding literacy than the one it replaces. It is also, I think, a more honest description of what universities were always supposed to be teaching, and a more compelling reason to keep teaching it.

References

1. Cheng, A., Calhoun, A. and Reedy, G. (2025) Artificial intelligence-assisted academic writing: recommendations for ethical use. *Advances in Simulation*, 10(22). Available at: https://doi.org/10.1186/s41077-025-00350-6.
2. Floridi, L. (2023) AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, 36(15). Available at: https://doi.org/10.1007/s13347-023-00621-y.
3. Floridi, L. (2025) Distant Writing: Literary Production in the Age of Artificial Intelligence. *Minds and Machines*, 35(30). Available at: https://doi.org/10.1007/s11023-025-09732-1.
4. Heaven, W.D. (2023) *ChatGPT is going to change education, not destroy it*, *MIT Technology Review*. Available at: https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/ (Accessed: April 5, 2026).
5. Hosseini, M., Resnik, D.B. and Holmes, K. (2023) The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*, 19(4), 449–465. Available at: https://doi.org/10.1177/17470161231180449.
6. Kaebnick, G.E. *et al.* (2023) Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing. *Medicine, Health Care, and Philosophy*, 26(4). Available at: https://doi.org/10.1007/s11019-023-10176-6.
7. Kim, J. *et al.* (2025) Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30(1). Available at: https://doi.org/10.1007/s10639-024-12878-7.
8. Kim, J. *et al.* (2026) Students-Generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*, 38(1). Available at: https://doi.org/10.1007/s12528-025-09444-6.
9. Leung, T.I. *et al.* (2023) Best Practices for Using AI Tools as an Author, Peer Reviewer, or Editor. *Journal of Medical Internet Research*, 25(e51584). Available at: https://doi.org/10.2196/51584.

10. Lin, Z. (2024) Techniques for supercharging academic writing with generative AI. *Nature Biomedical Engineering*, 9(4). Available at: https://doi.org/10.1038/s41551-024-01185-8.
11. Nguyen, A. *et al.* (2024) Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(5). Available at: https://doi.org/10.1080/03075079.2024.2323593.
12. Perkins, M. and Roe, J. (2024) Academic publisher guidelines on AI usage: A ChatGPT supported thematic analysis. *F1000Research*, 12(1398). Available at: https://doi.org/10.12688/f1000research.142411.2.
13. Pinker, S. (2014) *The sense of style: the thinking person's guide to writing in the 21st century*. New York, Viking.
14. The BSMT by Gianluca Gazzoli (2025) *Tutto quello che devi sapere sull'intelligenza artificiale! Prof. Luciano Floridi passa dal BSMT!* Available at: https://www.youtube.com/watch?v=q2VCtS4p1yQ (Accessed: April 18, 2026).
15. Villasenor, J. (2025) AI has rendered traditional writing skills obsolete. Education needs to adapt. *Brookings*, 30 May. Available at: https://www.brookings.edu/articles/ai-has-rendered-traditional-writing-skills-obsolete-education-needs-to-adapt/ (Accessed: April 3, 2026).

Stefano Adamo
Univerzitet u Banjoj Luci
Filološki fakultet
Katedra za romanistiku

OD PISARA DO ARHITEKTE

REFLEKSIJA PRAKTIČARA O PODUČAVANJU AKADEMSKOG PISANJA U DOBA VELIKIH JEZIČKIH MODELA

Rezime

Ovaj rad polazi od zapažanja iz učionice: studenti na predmetu akademsko pisanje počeli su sredinom semestra predavati neočekivano tečno napisane preglede literature, npr. pažljivo strukturisanu prozu, primjereno ograđenu ali, pri pažljivijem čitanju, djelimično izmišljenu. To zapažanje ukazuje na strukturni problem. Decenijama su nastavnici tretirali tečnost pisanja kao upotrebljiv pokazatelj intelektualnog angažmana, a onda je generativna vještačka inteligencija razbila taj pokazatelj, ne poboljšavajući studentsko razmišljenje, nego čineći uglađenu akademsku prozu dostupnom bez njega. Centralna tvrdnja ovog rada jeste da vještačka inteligencija pojača-

va postojeću kompetenciju, ali ne nadoknađuje njeno odsustvo. Studenti koji već znaju kritički čitati, formulisati argument i procijeniti validnost izvora koriste VI produktivno, a oni koji to ne znaju bivaju dovedeni u situaciju da samo izgleda kao da znaju. Na toj osnovi, rad predlaže konkretan trofazni radni tok izgrađen oko strukturisanog pisanja promptova, distribuisanog istraživanja putem više VI alata i progresivnog razgovora u namjenski izgrađenom VI okruženju. Radni tok osmišljen je da podučava pregled literature kao čin oblikovanja, a ne prepisivanja. Šira implikacija je pedagoška: ako kurs više ne može vrednovati pisanje kao gotov proizvod, mora vrednovati odluke koje to pisanje proizvede. Podučavati akademsko pisanje u doba velikih jezičkih modela znači podučavati prosuđivanje.

▸***Ključne riječi:*** akademsko pisanje, generativna vještačka inteligencija, veliki jezički modeli, pregled literature, pedagogija, VI pismenost, autorstvo, visoko obrazovanje.