Human–LLM Synergy in Higher Education Publishing: Two ChatGPT Use Cases within Editorial Pipelines

Gianluca Pavani[1],
[1] Università degli Studi di Roma Tor Vergata, Italia - gianlpav94@gmail.com
Edra S.p.A., Italia - g.pavani@lswr.it

ABSTRACT (ITALIANO)
Il presente contributo intende analizzare l'utilizzo di Large Language Models (LLM), con particolare riferimento a ChatGPT, nell'ambito dell'editoria scientifica e accademica. Sebbene il ruolo dell'intelligenza artificiale generativa nella scrittura creativa e nella narrativa sia stato ampiamente esplorato, la sua integrazione nei processi di scrittura scientifica ha ricevuto, ad oggi, una attenzione significativamente più limitata. Gran parte del dibattito attuale si concentra su implicazioni di natura etica e sul possibile indebolimento del rigore scientifico e autoriale. Questo studio adotta un approccio qualitativo per esaminare in che modo i LLM possano essere concretamente integrati nei flussi editoriali, presentando due casi applicativi sviluppati nel contesto dell'editoria universitaria. L'obiettivo è individuare metodologie operative e buone pratiche emergenti che favoriscano un utilizzo responsabile ed efficiente dell'intelligenza artificiale nella produzione di contenuti accademici.
Parole chiave: Large Language Models (LLMs); ChatGPT; scrittura scientifica; editoria universitaria

ABSTRACT (ENGLISH)
Human–LLM Synergy in Higher Education Publishing: Two ChatGPT Use Cases within Editorial Pipelines. This paper investigates the use of Large Language Models (LLMs), with a focus on ChatGPT, in the field of scientific and academic publishing. While the role of generative AI in creative and fictional writing has been extensively explored, its integration into scientific writing processes has received comparatively limited attention. Much of the existing discourse remains centered on ethical implications and the possible weakening of scientific rigor and authorial credibility. In contrast, this study takes a qualitative approach to examine how LLMs can be practically embedded in editorial workflows, presenting two real-world applications within higher education publishing. The objective is to identify concrete methodologies and emerging best practices that support the responsible and efficient use of AI in the production of academic content.
Keywords: Large Language Models (LLMs); ChatGPT; scientific writing; academic publishing

1. INTRODUCTION
The spread of generative artificial intelligence (AI) systems, particularly OpenAI's well-known ChatGPT, has sharply brought issues of potential uses, regulations, and risks into the educational landscape. The wide range of these crucial topics encompasses all stakeholders: authors, publishers, academic and school settings, professors, researchers, and students. This paper aims to investigate potential pipelines for implementing generative tools. The two case studies presented were conducted in collaboration with *Edra*, an Italian medical-scientific publishing house, and the discussed results will be considered as possible good practices for generative AI tools in the editorial production chain of educational content. Finally, the paper is also intended as a study on the AI's ability in scientific writing, a much more complex field than creative and fiction writing. Generative systems, often seen as unreliable facilitators prone to hallucinations, needs to be examined from a multi-level and interdisciplinary perspective, with an awareness that the effort of stakeholders should be directed towards proper implementation and regulation, rather than ineffectual limitations.

2. WRITING DRAFTS FROM NOTES: PROMPT-BASED APPROACH TO CHATGPT
The case study under examination pertains to an editorial project for the medical-scientific catalog of the Italian publishing house Edra. This project aims to publish two volumes of university textbooks on Biotechnology. The editorial initiative is part of a larger project comprising six volumes on Biomedicine, scheduled for publication by 2026. The first two volumes, "Principles of Biotechnology" and "Biotechnology Drugs," are anticipated to be released by early 2025 and will target an audience of approx. 1300-1500 students, with about two-thirds corresponding to the first volume, which is intended for the Bachelor's degree cohort.

The case study involved the application of ChatGPT, in versions 4 and 4o, for building the draft of the first two volumes of Biotechnology. The first manual is structured into eleven chapters and second one into twelve, totaling approximately three hundred and twenty printed pages each. OpenAI's LLM was utilized to generate drafts for twenty-one chapters. The editorial project involves a scientific board, composed of the headline authors, and coordinated scientifically by three external editors. Both the editors and the authors serve as Associate Professors or Full Professors at Biotechnology Faculties in the Universities of Milan (Università di Milano Statale, Università di Milano-Bicocca, Università Vita-Salute San Raffaele).
Drafting process using ChatGPT began with topics outlined in PPTX files, which comprised over five hundred slides in total. These PowerPoint files were created by the headline authors and their assistants as teaching support in courses, subsequently collected and selected by the scientific editors according to the editorial project. Consequently, the provided material, as a source of interpreted literature, is considered scientifically validated from the outset. The use of ChatGPT in versions GPT-4 and GPT-4o can be described as an extension of the content from all over five hundred slides, a process referred to by the system as "Transforming Slide Concepts". ChatGPT was employed throughout the entire process of drafting, expanding the textual content present in the PPTX files. The available slide material typically consists of unstructured texts, bullet points, images, charts, and quotes, which were transformed into extended and argumentative text. Furthermore, ChatGPT was tasked with providing a logical sequence and coherent indexing of the generated paragraphs, while maintaining existing structures from the slides (e.g., bullet and numbered lists) and extending their contents. GPT-4 and 4o were also employed to create captions for some images that lacked them.
The entire draft production process (e.g., Figure 1) was based on two primary criteria: the presence of external forms of scientific validation and the adoption of appropriate prompting strategies.

1. Forms of Scientific Validation: given the use of generative AI tools in drafting scientific texts for educational purposes, primary importance was placed on the presence of sufficient validation forms at both ends of the process. The first form of validation is present from the inception of the case study. The prompts exclusively included texts, data, graphs, images, and quotes from the provided PPTX files. As previously mentioned, the slides supplied by the authors were an integral part of the university teaching material, considered as sources of interpreted literature and thus validated upstream. These slides were deemed consistent with the editorial project during the material collection by the coordinating editors. The coordinating editors also handle the initial review of the scientific content. This phase is preceded by a structural check of the drafts by the catalog editor and followed by submission to the authors for final revisions, including any necessary cuts and additions. The outcome will be the second draft, formatted in Word, which will undergo another review by the coordinating editors for final approval.

2. Prompting Strategies: the efficiency of utilizing LLMs (Large Language Models) such as ChatGPT is inextricably linked to the proper construction of prompts, particularly in the context of models that perform reasoning tasks (Mialon et al., 2023). During the process of expanding unstructured scientific text into extended and argumentative prose, having prompting strategies is considered *conditio sine qua non* for obtaining satisfactory outputs. The strategies adopted can be categorized as follows:
    - Extent Measurement: inclusion of specifics regarding the desired length for text extension provided in the prompt. ("*Extend the text provided in the prompt to no less than (n. of tokens)*," with the highest efficiency achieved within a range of two hundred to five hundred tokens. In this phase, prompts were mainly zero-shot.
    - Linguistic Types of the Text to be Generated: prompts included instructions on the linguistic register and terminology to be used. ("*Use formal language appropriate for university textbooks*", "*Utilize terminology belonging to the technical jargon of the subject – Biotechnology*"). For this task, one-shot prompts have been largely used.
    - Structure of the Text to be Generated: essential for obtaining the desired textual form. ("*Transform into extended and argumentative text, divide into titled paragraphs*", "*Insert bullet and numbered lists where necessary and maintain those already present*"). This phase required mostly few-shot prompts.
    - Variable Prompts for Targeted Corrections: inserted as needed to correct parts of the generation, significantly facilitating the editing phase. ("*Avoid redundancies in the conclusions*").

The extension process, started at the end of April 2024, produced twenty-one AI-generated drafts of the first two volumes of Biotechnology by the end of August 2024. To date, all drafts generated with ChatGPT have been considered scientifically and editorially coherent from the first review by the editors.
From the application of ChatGPT for the generation of drafts in the medical-scientific editorial context, the following observations were made:

- Constant attention was directed towards some of the major issues related to the generation of scientific texts, namely the potential occurrence of hallucinations and biases in the outputs. Neither during the control phase in editing nor from the scientific evaluation by the editors did any hallucinations or biased generations emerge.
- ChatGPT proved effective in the extent and quality of text expansion, in the creation of ad hoc captions for images, including complex ones. Additionally, it provided correct indexing and titling of paragraphs, helping to build the structure of subsections, paragraphs, and subparagraphs (the overarching chapter structure was known from the start).
- ChatGPT capabilities allowed for minimizing the editing phase, which primarily focused on correcting redundancies and occasional typos.
- By providing scientifically validated material in the prompt, ChatGPT transformed the slide content into coherent text with logical sequencing and a remarkable level of quality and originality. Furthermore, through AI, it was possible to contextualize and integrate quotes, data, documentary content, and images that were otherwise detached from the rest of the content in the slides.

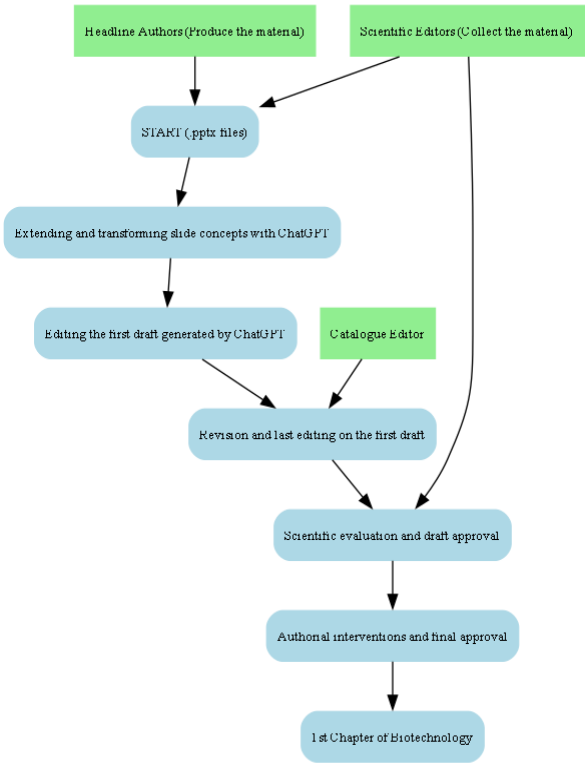

Figure 1. Workflow and validation steps

3. DOMAIN-GROUNDED, PROMPT-BASED CUSTOMIZATION AND EMBEDDED DOCUMENT CONTEXT VIA CHATGPT API

The second case study presented here illustrates an alternative implementation – currently ongoing – of a Large Language Model within editorial workflows. In contrast to the first project, which focused on extending and reshaping unstructured source material, this initiative aims to distill content from extensive textual sources for the purpose of generating structured drafts. Final output will consist of a concise revision study guide book designed for second and third-year medical students enrolled in the Medical Physiology course. The project involves collaboration among a subject-matter expert responsible for curating content and defining the indexing structure, as well as editors and catalog editors from the

publishing house. Final publication is scheduled for late 2025, pending further refinement of the editorial project.
Production process relied on both open-access materials and proprietary textbooks and study guides from the Edra catalog. Given the need to compress and restructure large bodies of content, and in light of the token limitations of direct prompt-based interaction with the chatbot interface, the workflow adopted a more technically structured approach using the ChatGPT API, Python scripting, and document storage functionalities. The technical workflow and editorial integration followed the key stages detailed as follows:

1. Model and Temperature Control: the language models (GPT-4 and GPT-4o) were accessed via API and configured with Python to ensure consistent, accurate, and context-aligned summarization. After some testing, the temperature parameter was adjusted within a range of 0.2 and 0.5. Lower values (0.2–0.3) were used for dense, technical chapters ensuring terminological precision and minimal creative deviation. Slightly higher values (0.4–0.5) were tested when summarization required more abstraction and paraphrasing.
This regulation allowed the model to maintain both semantic fidelity and stylistic consistency with the domain-specific content.
2. File Upload and Contextual Anchoring: source documents, such as university-level textbooks, were uploaded as TXT files to the ChatGPT file storage via API. Each document was then referenced in the prompt using its file ID in Python, providing the model with grounding in domain-specific terminology and disciplinary structure. These files acted not only as the objects to be summarized, but also as semantic and stylistic context, anchoring the generation in the established discourse of medical physiology.
3. Chunking Strategy: files were segmented to manage long documents and align with token window limitations. Each chunk was sized at approximately 1500-2000 tokens, equivalent to 1000–1600 words depending on the source format. This granularity allowed sufficient context retention while keeping the input within manageable token limits for API calls, showing consistency and completeness in summarization, with balance between context and focus.
4. Target Output Length and Compression Ratio: each input chunk was summarized into approximately 200–400 tokens, which equates to a compression ratio of 5:1 to 8:1, depending on the source density and repetition. This compression level ensured that all relevant concepts were preserved in a concise, revision-oriented form, avoiding excessive abstraction or loss of terminological accuracy.
5. Persona, Prompting and Content Constraints: the prompt consistently assigned the model the role of a senior ghostwriter with expertise in scientific and medical writing, specialized in Medical Physiology. Beyond directing the model toward coherent disciplinary tone with the persona, each prompt included specific instructions to maintain alignment with the source material. These prompts enforced:

- Terminological precision, avoiding simplification of medical vocabulary.
- Tone appropriate for academic writing, neutral and informative rather than creative or colloquial.
- Structural mirroring of the source material, when applicable (e.g., maintaining topic subdivisions and preserving sequential logic).

The prompt reinforced that the summaries should serve as revision tools, assuming that the reader had prior exposure to the full content via textbooks and lectures.

6. Human Oversight and Indexing: the output underwent a manual editorial phase by expert and editors, ensuring fidelity to scientific and medical standards. This phase included refining terminology and phrasing, eliminating redundancies occurring across adjacent chunks (use of ChatGPT was included for this task) and validating the factual accuracy of complex physiological descriptions. Subsequently, all summaries were indexed by the editorial team of the publishing house, according to the structure developed by the collaboration between expert and university-affiliated library partner.

Between July and November 2024, the process resulted in the completion and editorial approval of draft versions for twelve chapters of the Medical Physiology study guide book. This timeline primarily reflects the iterative nature of editorial validation and content refinement. Since all processing is handled via the OpenAI API, the workflow itself remains technically lightweight and the infrastructure requirements minimal, making this approach scalable and replicable even in small editorial settings with limited resources.

4. DISCUSSION

It is important to preface the following discussion of the two case studies with the observation that it is primarily grounded in a coherent interpretation of the results obtained, given the previously mentioned scarcity of literature on comparable use cases or studies.

Firstly, from a general observation, ChatGPT demonstrates remarkable effectiveness in writing activities within the editorial workflow of a university textbook. Moreover, given the generation of texts pertaining to a well-defined domain, the study also examined ChatGPT's behavior in the complex task of scientific writing. The application of ChatGPT —and LLMs in general— in the field of scientific writing is subject to significant controversy. The main ambiguities revolve around scientific reliability, respect for authorship, and copyright. While it is true that these issues are widely discussed in the literature, many of the existing studies have stemmed from considerations on cases different from the ones presented here, such as the growing use of generative AI as a ghostwriter in scientific articles, papers, and literary reviews. This phenomenon, particularly in the medical-scientific field, is perceived as a threat to the integrity of academic ethics (Khalifa, 2023); the reason lies in the partly compromised principle of authors' agency, as well as in the risk of plagiarism, which is notably difficult to detect when GPT is tasked with generating scientific content.

However, these risks, though undoubtedly significant, seem to be particularly relevant in two circumstances. The first involves generation from scratch, or more precisely, generation based solely on the statistical-probabilistic criteria applied to GPT's memory. Although improvements have been made in this regard with the latest versions, ChatGPT is not designed to provide precise bibliographical references and struggles to distinguish Open Source literature, moreover in previous versions it provided wrong PMID numbers (Alkaissi & McFarlane, 2023). Additionally, there is the risk of potential hallucinations, such as fictitious references (Babl & Babl, 2023). The second circumstance involves the input of external material into the prompt to summarize existing literature on a specific topic (Salvagno et al., 2023).

When implementing ChatGPT within the first editorial project, all the major known issues were taken into account, and the factor that enabled justified skepticism to be overcome was the inherent characteristics of the project itself. The principle of authorship, or authors' agency if preferred, is respected insofar as the validity of the draft produced by ChatGPT—scientifically, conceptually, and terminologically—has been verified. It is the authors' additions and modifications that ultimately shape the final version of the text. The issue of bias and hallucinations is strongly mitigated by the use of sources that are considered validated at the outset (albeit in an unstructured form) and, above all, through the adoption of various steps to validate scientific rigor by qualified personnel. It is worth emphasizing that all drafts produced were approved by the scientific board, as no hallucinations or significant errors emerged in the generation of scientific content.

While this first case demonstrates the potential of LLMs for expanding loosely structured content through flexible, interface-based prompting, it could also revealed certain limitations in terms of replicability and scalability on domain control, consistency, and contextual fidelity, particularly when working with larger textual contents. To explore how these challenges might be addressed through a more structured and domain-grounded setup, the second case study shifts focus toward the summarization of complex source materials via API orchestration. This approach prioritizes semantic alignment and editorial control, offering a technically more sophisticated, albeit narrower, implementation of LLM-assisted content production.

In the second case of use, rather than on a fine-tuning, the system relied on a domain-specific adaptation pipeline based on the ChatGPT API and documents as contextual anchors within the generation process; this technique can be described as a lightweight domain adaptation strategy, on top of a general-purpose LLM via API-level configuration. While not altering the underlying model architecture or parameters, the system effectively simulated domain-specific competence by combining embedded content references with a document-linked prompt orchestration structure. In a way, a dynamic and semi-structured form of knowledge grounding – clearly distinct from Retrieval-Augmented Generation (RAG) – yet could be functionally comparable in terms of enhancing relevance and coherence within a predefined domain.

The result so far seems to be a controllable and reproducible output pipeline tailored to the editorial aim of producing drafts of high-utility texts for study guide books.

Both use cases can lead to broader observations about the content generated by ChatGPT: different types of writing—on one hand generalist, creative, and fictional; on the other, scientific, in-depth and rigorous—inevitably do not suffer from the same critical issues. The use of LLMs in creative writing is often regarded as a potential threat to the quality and variety of content, and the greatest risks are particularly associated with its use during school-age years, when writing skills are still being developed (Shidiq, 2023). Indeed,

while generative AI can be credited with a certain degree of originality (which is undoubtedly improving in the latest versions of the models), widespread use of these tools would sooner or later lead to redundant content, both in terminological and syntactic form and in substance. As mentioned, this application of AI has been primarily studied in other contexts, and the majority of the literature on the topic focuses on ethical concerns (Otmar et al., 2024). Only a few studies have attempted to explore alternative forms of collaboration aimed at positively impacting the performance of both parties —humans and LLMs— involved (Chackrabarty, Laban & Wu, 2025). Research concerning the use of LLMs in scientific writing—particularly in the medical field—has almost exclusively addressed the drafting of journal articles or literature reviews (Ahn, 2024), rather than editorial practices related to the publication of academic textbooks.

However, based on the findings of this editorial project, their use in the context of scientific writing seems to lead to a different paradigm. Scientific knowledge, particularly that aimed at academic teaching, is inherently more fixed, despite its constant evolution. In a way, scientific writing allows for greater flexibility in form, provided that the essential requirement of rigor is upheld. In this sense, the use of LLMs could be seen as a tool to facilitate the transmission of scientific knowledge, as they are highly effective at reinterpreting its form. The case presented here constitutes an example, as ChatGPT was not used to produce significant scientific insights but rather to structure non-cohesive and non-argumentative texts, leveraging its Natural Language Processing (NLP) capabilities. Through AI, it was possible to transform unstructured scientific sources from interpreted literature, such as slides used in university lectures, into a different textual form in significantly less time than was previously required.

In other words, it could be argued that LLMs may represent highly valuable tools in the future for reworking and disseminating scientific knowledge (Markowitz, 2024), or contribute to making scientific communication more accessible, as has been particularly noted in the use of small models subjected to fine-tuning (Jiang, Shi & Luo, 2024). From an editorial perspective, it could translate into greater productivity in publishing and more rapidly updatable catalogs. Such benefits represent an undeniable advantage in presenting scientific knowledge that is aligned with the state of the art, which is fundamental in rapidly evolving fields, such as the medical-scientific one, as well as in more promptly addressing the needs of students and, more generally, the academic environment.

Despite the divergent technical setups and editorial aims of the two case studies, they share two fundamental commonalities. In both cases, the textual material served a dual function; it operated simultaneously as the target of generation (i.e., the content to be transformed or summarized) and as a source of contextual and domain-specific grounding. Moreover, both projects were characterized by a structured collaboration between human academic and editorial figures and the language model, with the latter operating not in isolation but as a component within a carefully supervised pipeline.

These findings suggest that, rather than striving for full automation, the human–LLM synergy can offer a more promising and sustainable path forward for integrating generative models into academic publishing workflows.

5. CONCLUSIONS

The results emerging from both case studies support a broader reflection on the evolving role of generative language models within editorial and publishing ecosystems. The main challenges associated with these systems have often been treated as limitations to be highlighted rather than as opportunities for exploring potential solutions. Rather than replacing human expertise, the integration of LLMs into these pipelines points toward a symbiotic model of collaboration, in which the machine enhances human capabilities without displacing them. This human–LLM synergy enables a division of labor that is both effective and sustainable: the model contributes scale and consistency, while human editors and experts ensure semantic precision, disciplinary relevance, and alignment with pedagogical intent. Such an approach reframes generative AI not as a tool of full automation, but as a cooperative agent embedded in domain-specific processes, aligning with emerging paradigms of human-in-the-loop knowledge production, where credibility and scientific rigor remain central. If we consider the balance between efficiency, cost-effectiveness, time constraints, and reliability as crucial for the application of AI systems in higher education publishing contexts, collaborative and symbiotic frameworks may enhance the domain relevance and the trust in the outputs—key factors for a broader acceptance, and consequently applicability, of LLMs in this field.

REFERENCES

Ahn, S. (2024). The transformative impact of large language models on medical writing and publishing: current applications, challenges and future directions, Korean J Physiol Pharmacol 2024;28(5):393-401.
https://doi.org/10.4196/kjpp.2024.28.5.393

Alkaissi, U., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing, *Cureus* 15(2): e35179.
DOI 10.7759/cureus.35179

Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract?. Emergency Medicine Australasia, 35: 809-811.
https://doi.org/10.1111/1742-6723.14233

Chakrabarty, T., Laban, P., & Wu, C. S. (2025). Can AI writing be salvaged? Mitigating Idiosyncrasies and Improving Human-AI Alignment in the Writing Process through Edits, in Yokohama' 25: ACM (Association of Computing Machinery) CHIconference on Human Factors in Computing Systems, April 26–May 1, 2025, Yokohama, Japan. ACM, New York, NY, USA.
https://doi.org/10.48550/arXiv.2409.14509

Jiang, G., Shi, X., & Luo, Q. (2024). LLM-Collaboration on Automatic Science Journalism for the General Audience.
https://doi.org/10.48550/arXiv.2407.09756

Khalifa, A. A. (2023). Does the use of AI tools in academic publishing represent a new form of ghost authorship? *Croatian Medical Journal*, 64:292-4.
https://doi.org/10.3325/cmj.2023.64.292

Markowitz, M. M. (2024). From Complexity to Clarity: How AI Enhances Perceptions of Scientists and the Public's Understanding of Science. *PNAS Nexus*.
https://doi.org/10.48550/arXiv.2405.00706

Mialon, G., et al. (2023). Augmented Language Models: a Survey, in Machine Learning Research, June 2023.
https://doi.org/10.48550/arXiv.2302.07842

Otmar, R., et al. (2024). Ethics and the use of generative AI in professional editing.
https://doi.org/10.1007/s43681-024-00521-7

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing?, *Critical Care* (2023) 27:75.
https://doi.org/10.1186/s13054-023-04380-2

Shidiq, M. (2023). The Use of Artificial Intelligence-Based Chat-Gpt and its Challenges for the World of Education; from the Viewpoint of the Development of Creative Writing Skills. *Proceedings of International Conference on Education, Society and Humanity*, Vol. 01 No. 01 (2023).
https://ejournal.unuja.ac.id/index.php/icesh