International Journal of
Cancer Care and Delivery

Conference Abstracts
Vol. 6, Issue Supplement 1, 2026

Leveraging Large Language Models for Intelligent Document Processing in Medical Journal Publishing

Pratik Hadkhale, BS[1,a]; Bipul Luitel, Phd.[1]; Suyog Kadel, Msc.[1]

[1] Binaytara, 3380 146th Pl SE, Bellevue, Washington, United States

Corresponding author
[a] Pratik Hadkhale, BS, Binaytara, 3380 146th Pl SE, Bellevue, Washington, United States
Email: pratik.it@binayfoundation.org

Keywords: Large Language Models, Document Processing, Journal Publishing, Natural Language Processing, Automated Typesetting

Submission received: 2025-09-18 / Accepted: 2026-01-07 / Published: 2026-01-25

https://doi.org/10.53876/001a.129609

Background

The academic publishing industry processes millions of research articles annually, with medical journals facing unique challenges due to complex formatting requirements, specialized terminology, and strict accuracy standards. Traditional document processing workflows in journal publishing involve multiple stages: manuscript submission, peer review, copyediting, typesetting, and final formatting for web publication. Each stage historically required significant human intervention, particularly the conversion from author-submitted formats (typically Microsoft Word or PDF) to structured, web-ready formats (HTML/XML).

Methods

Traditional approaches to converting submitted manuscripts from PDF and Word formats to web-ready HTML/XML involve significant manual effort, are error-prone, and incur substantial costs. Our system, deployed at Binaytara's journal platform, leverages Claude (Anthropic) to intelligently parse, structure, and format medical research articles while preserving semantic integrity and handling complex elements such as references, equations, and medical terminology.

Manuscripts submitted in Microsoft Word format through our journal portal undergo an automated processing pipeline. First, submissions are converted to PDF via the Convert API to facilitate inline annotations during peer review. After acceptance through peer review, the original Word file is processed using Markitdown to extract the text content. This content is then processed by a large language model, which reconstructs the article in structured JSON format, which consists of defined sections (e.g. Abstract, Introduction, Methods, Results, Discussion, References), formats inline references, equations, and headings, and resolves edge cases that challenge heuristic or rule-based approaches, particularly in reference formatting. The structured JSON format is then used to create formatted HTML, PDF and JATS XML versions of the article.

Results

The new platform achieved a 95% reduction in processing costs compared to traditional third-party services and lowered total management and processing costs by over 60%. Efficiency gains included immediate post-review formatting, elimination of external vendor dependencies, and reduced copyeditor workload, enabling greater focus on content quality. Key benefits comprised a cost-effective publishing workflow, faster publication turnaround, improved consistency

and control over article structure, and enhanced editorial capacity.

Conclusion

The integration of LLMs into our journal platform has proven to be a transformational upgrade to our publication workflow. By automating typesetting and reference formatting, we have drastically cut costs, improved efficiency, and reduced editorial overhead. This demonstrates the practical value of AI in scholarly publishing and opens avenues for further innovations in digital content processing.

	Binaytara Journal Portal	Third-party Platform
Processing cost (per article) • **Platform fees** • **AI API fees** • **Conversion API**	$1	$30
Total articles published (per year)	100	100
Journal management cost • **Journal management's expenses**	$500	$1,000
Five year management cost	$30,000	$60,000
Five year processing cost	$500	$15,000
Total Cost	$30,500	$75,000
Cost Savings over 5 years	**$44,500 (~60%)**	

Cost comparison between Binaytara Journal Portal and Third-party Platform

© 2021–2026 Binaytara. All rights reserved. Articles published in the International Journal of Cancer Care and Delivery are open access articles distributed under the terms of the Creative Commons Attribution-ShareAlike (CC BY-SA) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited and derivative works are shared under the same license. | E-ISSN: 2770-3533