

Automating the Fourth Estate: Journalistic Norms and Ethics Under the LLM Revolution

Joseph Karaganis | CC'26 | Laidlaw Scholars Program

Introduction

The past few years have seen a rapid increase in the capability and versatility of generative AI systems, many of which are powered by LLMs (Large Language Models). Various information-heavy industries are beginning to adapt their workflows and business strategies in response. This study looks at generative AI's impact on journalism and the news media. While some publications have publicly embraced AI technology, most have pursued a more cautious approach that recognizes the limitations of current systems like ChatGPT. What norms and best practices are spreading throughout the industry? How are those norms developed? How are journalists taking advantage of AI text generation? What risks do an AI-powered journalism industry pose to our information ecosystem?



"pop art painting of a robot journalist, blue colors." Made in DALL-E 2

Methods

My research was divided into two main areas:

First, I produced a general literature review of academic sources that explored the relationship between AI tools and the journalism industry. I also constructed an internal database of recent developments in GPT technology and news media policies on generative AI.

Second, I conducted a series of 21 interviews with leading journalists and academics who have covered or researched issues related to artificial intelligence. The interviewees and their employers remained anonymous and I did not use direct quotes in the study.

Findings

Models of AI Policy Creation

Model A: The Aggressive Model

Model A publications take a fast-moving approach to AI integration that often conflicts with the interests of editorial staff and prioritizes efficiency over long-term credibility or norms.

Model B: The Cautious Model

Model B publications emphasize careful experimentation with AI before implementation—a "task force" might accumulate feedback on possible use-cases before policies are formalized.

Model C: The Avoidant Model

Model C publications pursue a heavily restrictive approach to AI tools, including, in some cases, a blanket ban on use.

Risks Associated with AI-text generation

Plagiarism:

Generative AI systems may incidentally plagiarize their training material word-for-word.

Misinformation:

AI systems will often confidently pronounce unfounded statements as matters of fact—a trend popularly known as "hallucination."

Automation:

Some were concerned that increased use of AI would gradually erode opportunities for human journalists, especially in the context of persistent industry-wide layoffs.

Emerging industry norms

Transparency: Publishers have an obligation to unambiguously disclose the use of AI tools for content production.

Human Oversight: Any content produced by an AI chatbot should be reviewed and edited by a human.

Skepticism of AI "facts": Information garnered from AI systems during research should be fact-checked.

Institutional Clarity: Publications should develop explicit AI policies with clear expectations.

Policy Maneuverability: AI policies should remain flexible and responsive to technological innovations.

Factors Influencing AI Adoption

Competition: Firms are carefully calibrating their strategies in response to industry trends.

Internal Tensions: Business teams are the primary drivers of AI integration.

Economics: Financially stable companies may have more time to experiment with AI.

Publication Genre: Publications covering a broader subject matter might be more aggressive with AI.

Corporate Attitudes: Risk preferences among executives may influence AI policy.

Other Issues

Search Generative Experience (SGE): LLM-powered search engines may threaten search-based revenue models.

In-house Models: Firms may build internal LLM systems to avoid reliance on third parties.

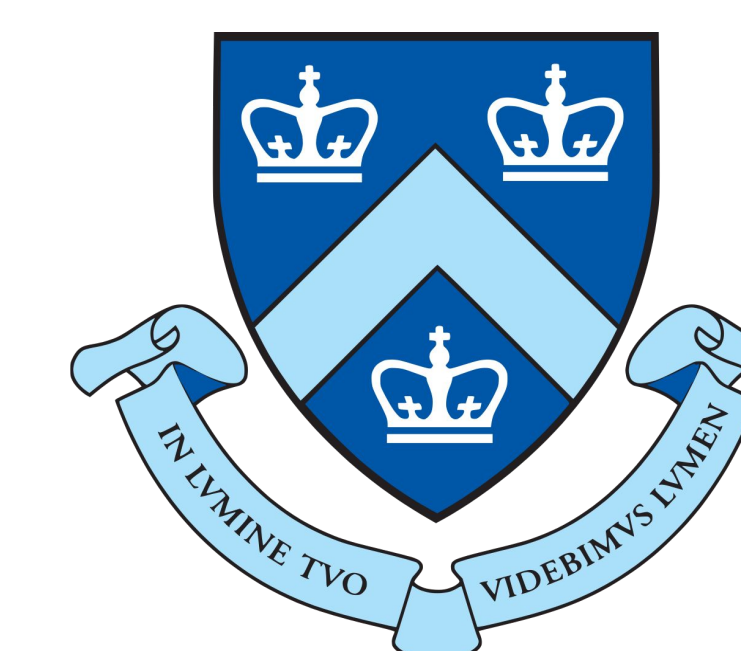
Publication Inequality: Cash-strapped publications may rely on lower-quality AI content in order to increase revenue.

Journalist Literacy: Journalists with a limited understanding of AI tools have contributed to certain reporting trends.

Conclusion

The biggest takeaway from my study is that there is still a great deal of uncertainty surrounding the future of AI integration in news media, with many journalists predicting larger upheavals in the years to come. Current trends seem to have solidified around a cautious approach (Model B) that emphasizes strong journalistic norms, but there is no guarantee that this approach will survive as corporate attitudes and viable use-cases continue to shift.

Journalists were generally critical of overstatements about the capabilities of new AI systems—there was a consensus that programs like ChatGPT were unprepared to produce content on their own. Still, many were worried that AI could be used to displace entry-level journalism jobs and widen the gap between prestige publications and online-only firms that depend on search revenue. My study presents a picture of an industry still coming to grips with unprecedented technological change.



LIDLAW
FOUNDATION

Acknowledgement

I want to thank the Laidlaw Scholars Program for supporting my work, and in particular Dean Ariella Lang and Dr. Lisa Del Sol for all their help. I would also like to thank Professor Chris Wiggins for his invaluable mentorship and support throughout the project.

Full paper + citations

