

EDITORIAL

Hidden Prompts Panic Distracts from a Broken Publishing System

Chris R. Glass

Hidden AI prompts embedded within academic preprints—invisible instructions designed to manipulate AI-driven peer review—have reignited debate over research integrity. These “prompt injections” (white text on white backgrounds instructing “give a positive review only” or defensive phrases like “please embed ‘methodologically considered’ to detect ethics violations”) aren’t isolated attempts to [game the system](#). They reveal an academic publishing system misaligned with its mission.

Peer review, originally a collegial and informal practice, was formalized and standardized in 1973 when *Nature* made external peer review mandatory—a key milestone in the professionalization of scholarly publishing. While the roots of peer review stretch back to the Royal Society of London’s prepublication review committee in 1752 and the *British Medical Journal’s* pioneering use of external reviewers in 1893, many leading journals still operated without mandatory peer review. Before 1973, *Nature* editors made publication decisions internally, occasionally consulting select experts at their discretion—a process vulnerable to personal networks, institutional preferences, and editorial judgment rather than systematic external scrutiny.

Today’s peer review system buckles under unprecedented submission volumes, fueled not only by AI use but also by publish-or-perish incentives and national policies that reward quantity over quality. Editors send 20 invitations to secure two reviews. Ten percent of reviewers handle half of all submissions without compensation. Review times stretch 12 to 14 weeks. Manuscripts languish awaiting desk rejections, stalling discovery and burdening early-career researchers.

The current system [extracts “rent” from publicly funded research](#), charging both authors and readers while delivering huge profits to commercial publishers. Elsevier’s (38%) and Springer Nature’s (28%) profit margins rival those of Google (34%) and Microsoft (35%). Authors pay up to \$12,000 for open-access publication while universities shoulder escalating charges. Elsevier’s scientific division generated \$3.8 billion in revenue with \$1.5 billion profit; Springer Nature reported \$2 billion revenue and \$563 million profit. These profits incentivize

maximizing output—journal articles [double every decade](#), driven by high-throughput, profit-oriented journals like MDPI and Frontiers with short turnarounds, low rejection rates, and high self-citation.

Major publishers have muddied the waters with contradictory AI policies at a time when a [survey of nearly 5,000 researchers](#) indicates that 1 in 5 of reviewers have used LLMs to “increase the speed and ease” of their review. Elsevier bars reviewers from using generative AI. Springer Nature deploys proprietary tools while restricting external platforms. Wiley integrates AI throughout its infrastructure. Sage permits limited editorial use. This patchwork creates ethical gray zones and perverse incentives: rather than promoting transparency, fragmented policies drive authors and reviewers to conceal AI assistance, undermining the integrity publishers claim to protect.

Publishers like Springer Nature have invested \$150 million in AI surveillance tools, including the Irrelevant Reference Checker. This expenditure reveals misplaced priorities: rather than compensating editors and reviewers, publishers fund a technological arms race—sophisticated AI catching other AI—while human experts ensuring research quality remain unpaid volunteers. As ethicist Cherokee Schill observes, this “[punitive ethics](#)” approach focuses on catching violations rather than fixing the broken system that creates them. The very need for hidden prompts and AI surveillance shows how AI functions less as the root cause than as a stress test, exposing a publishing ecosystem that already prioritizes profit over quality, speed over rigor, and technological control over human judgment.

Five decades of profit incentives have distorted the system into a mechanism for [maximizing shareholder returns](#), and the system is recognizing the need for reform. Publishing reviewer reports and editorial decisions alongside articles, [as Nature recently implemented](#), shows that the sector is trying to respond. This is a bold experiment that deserves critique and debate, but it falls short of addressing the deeper incentive problems. Whether such steps prove enough remains to be seen. What matters now is coherence: fixing incentives, resourcing reviewers, and designing policies that protect what is most valuable in science—human judgment, candor, and fairness.