PROMPTING CHANGE WITH GENAI IN LARGE-SCALE DOCUMENT REVIEW: A REAL-WORLD STUDY ON GENAI

Robert Keeling, *Ray Mangum*, *Amy Hanke*, *& Alyssa Ogden*

INTRODUCTION

The evolution of document review in civil litigation has reached another milestone—one that invites a new partnership between humans and technology that promises gains in accuracy, efficiency, and scalability. Generative artificial intelligence (genAI) is remaking business and professional life at extraordinary speed,[1] and, within the legal sphere, nowhere is its impact more immediate—or the potential changes more dramatic—than document review.

Yet significant questions remain. When does genAI meaningfully improve document review? How should it be integrated into established workflows? How should its outputs be validated? As with prior technologies that are now widely accepted in eDiscovery, genAI's broader adoption will depend on demonstrated reliability and value supported by empirical data.

---

1. *See, e.g.*, Charlie Campbell et al., *The Architects of AI are Time's 2025 Person of the Year*, TIME (Dec. 11, 2025), https://time.com/7339685/person-of-the-year-2025-ai-architects/ [https://perma.cc/GX7T-MCLX].

Document review in civil litigation has undergone significant transformations before.[2] In the 1990s and earlier, review was largely manual: teams of attorneys examined thousands (and sometimes millions) of paper documents in warehouses over months. The early 2000s ushered in "eDiscovery," as digital document collections and keyword searching became central to review and production. In the 2010s, technology-assisted review (TAR) (a.k.a. "predictive coding") introduced more sophisticated tools based on machine learning—along with much debate about acceptable use, validation, and defensibility. With each evolution came novel issues, skepticism, and—ultimately—broad adoption.[3]

Today, the claim is that genAI can perform large-scale document review more effectively, economically, and defensibly than earlier approaches. The profession needs more empirical data to test that claim—to evaluate whether genAI truly delivers on its potential across varying types of legal matters and document sets. Among the key questions:

- Does genAI accurately and consistently predict relevant vs. non-relevant documents better than earlier technologies?
- Does genAI outperform or at least match human reviewers?
- Is genAI a defensible review methodology under legal discovery standards?
- Are there use cases where genAI does not perform well?
- What are the actual time and cost savings?
- How can lawyers best integrate genAI into existing workflows?
- What risks, if any, arise from using genAI in document review?

In this Article, we seek to answer some of these questions with empirical data, providing insight into the use of genAI to assist with document review in large-scale legal and regulatory matters.

For this case study, we worked with a client company and an eDiscovery vendor to select a 1,600-document sample from the client's prior, real-life legal matter. We then compared the relevance and issue coding of human reviewers to the coding decisions made by a market-leading genAI tool, Relativity's aiR for Review. The results varied depending on document type, but overall, the relevance review produced a high recall rate of 89% and a strong precision rate of 85%. The issue-coding results were more mixed, with accuracy often depending on the

---

2. *See, e.g.*, Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient than Exhaustive Manual Review*, 17 RICH. J.L. & TECH. 11 (2011); George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J.L. & TECH. 10 (2007).

3. *See generally* Charlie Hernandez, *Tech Shifts & the Law*, 48-AUG L.A. LAW 26 (July/Aug. 2025).

nuances of the specific issue. The testing also revealed certain limitations of the tool. Based on these results, this Article concludes with recommendations for potential genAI use cases within a typical document review workflow and discusses the potential cost savings, including the factors that may affect that calculus.

Part I of this Article describes the evolution of document review methodologies from manual review to TAR and concludes with the rise of genAI and its early impact on the legal profession. In Part II, we discuss the need for empirical data to evaluate the utility and defensibility of genAI tools in document review, and we describe our proof-of-concept study. In Part III, we offer an analysis of the study's results and provide our observations, learnings, and recommendations on how lawyers and their clients can best leverage genAI to enhance efficiency, improve accuracy, and make informed decisions about integrating those technologies into their existing document review workflows.