DistillerSR

How Pfizer Leveraged DistillerSR's Smart Evidence Extraction (SEE) to Cut SLR Data Extraction Time by 35%

35% faster extraction – SEE reduced time per study from 85 to 55 minutes, saving approximately 9.5 hours across 10 studies

Bounded AI reduces hallucination risk – context-aware architecture constrains outputs to the source document and extraction form, not general pre-training data

Human-in-the-loop is non-negotiable – every SEE suggestion requires active reviewer acceptance or rejection, preserving expert oversight and regulatory defensibility

Full audit trail built in – evidence linking and DistillerSR's auditable workflow ensure complete traceability from source document to extraction output

Copyright and privacy compliant – all data remained within the platform; no content shared with third-party AI vendors

> "The use of an integrated AI with HITL platform proved effective in enhancing efficiency and reducing the time burden of data extraction for COVID-19 SLR. These data suggest that there is value in using purpose-built AI/HITL platforms in reducing the data extraction burden in systematic literature reviews, contributing to more timely and cost-effective evidence synthesis on important public health topics."
>
> **– Said et al., Pfizer Ltd. / DistillerSR**

The Challenge: Data Extraction as a Bottleneck in Evidence Synthesis

Systematic literature reviews are foundational to pharmaceutical evidence generation – but the data extraction phase is consistently the most time-consuming and labor-intensive step in the process. For each included study, a trained reviewer must read the full text, locate relevant data points, and manually enter structured information into an extraction form. This process is repeated across every included study, by multiple reviewers, with adjudication required when answers conflict.

For a COVID-19 SLR examining the comparative vaccine effectiveness of different vaccines administered within the same respiratory virus season, the Pfizer team faced this familiar challenge at scale. With 19 studies meeting inclusion criteria, the team needed a faster, more efficient approach to extraction – one that could reduce time burden without compromising the accuracy or auditability of the outputs.

The Solution: DistillerSR's Smart Evidence Extraction (SEE)

The team conducted the SLR on the DistillerSR platform, registered in PROSPERO (CRD42025635806), and applied DistillerSR's Smart Evidence Extraction (SEE) feature during the data extraction phase.

SEE is a bounded Generative AI tool that operates directly within the DistillerSR workflow. Unlike general-purpose large language models, SEE is context-aware – its responses are constrained by the specific data extraction form and the content of the study PDF being reviewed, rather than drawing on broad pre-training data. This architecture is specifically designed to reduce hallucinations while maintaining the precision required for regulatory-grade evidence synthesis.

The team split the 19 included studies across two extraction methods: SEE with human-in-the-loop (HITL) validation was used for 10 studies, while 9 studies were extracted manually by a separate researcher. This parallel design enabled a direct, controlled comparison of accuracy and time efficiency between the two approaches.

How SEE Works in Practice

AI Suggestion: SEE reads the source PDF and suggests structured answers to pre-specified extraction questions, drawing only from the content of the document being reviewed

Evidence Highlighting: SEE highlights the supporting passage in the source document and provides attention maps, natural language explanations, and a confidence score for each suggested answer

Human-in-the-Loop Validation: Reviewers must actively accept or reject each suggestion – SEE does not auto-populate extraction forms, ensuring expert oversight at every step

Evidence Linking: Accepted answers are linked directly to the source passage, creating a transparent, traceable record for audit purposes

Copyright and Privacy Compliance: All data and references remained within the DistillerSR platform and were not shared with any third-party generative AI vendors

The Results: Substantial, Measurable Time Savings

The Results

35% Reduction in Extraction Time Per Study

The most significant outcome was a measurable reduction in the time required for data extraction. SEE-assisted extraction took an average of 55 minutes per study, compared to 85 minutes per study for manual extraction – a 35% reduction in time per study.

Across the 10 studies extracted using SEE, this translated to a total time saving of approximately 9.5 hours – a meaningful efficiency gain for a research team working under publication and submission timelines.

Accurate Capture of Structured and Project-Specific Data

SEE accurately captured standardised study characteristics including study design, population demographics, and other core data fields. When clearly reported in the publication, it also successfully identified project-specific data points such as COVID-19 variant information – demonstrating its ability to go beyond boilerplate extraction to support nuanced, therapy area-specific reviews.

The tool encountered occasional difficulties with complex results tables, which required manual capture. This finding reinforces the importance of the human-in-the-loop model: SEE accelerates the high-volume, structured elements of extraction while keeping expert reviewers in control of the most complex data.

Hallucination Mitigation Through Bounded AI

A key concern with generative AI in regulated research environments is the risk of hallucination – outputs that are plausible but factually incorrect. SEE addresses this through its context-aware architecture: rather than generating responses from general pre-training data, it focuses its outputs exclusively on the document under review, constrained by the extraction form. This approach demonstrably reduced hallucination risk while preserving the efficiency benefits of AI assistance.

Full Auditability and Regulatory Compliance

Because SEE is integrated into DistillerSR's auditable workflow, every extraction decision – whether accepted or rejected – is logged with a complete audit trail. The evidence linking feature connects each accepted answer to its source passage, providing the documentation required for regulatory submissions and peer-reviewed publication.

The Results: Substantial, Measurable Time Savings

The Pfizer COVID-19 SLR demonstrates that purpose-built, bounded AI with expert-in-the-loop validation can meaningfully reduce the time burden of data extraction in systematic literature reviews – without sacrificing accuracy, auditability, or compliance.

For pharmaceutical teams conducting high-volume SLRs across multiple therapy areas and submission deadlines, the implications are significant: faster extraction cycles, reduced reviewer burden, and a defensible, traceable evidence base that meets the standards of both peer reviewers and regulatory agencies.

About Pfizer

Pfizer Ltd. is the UK affiliate of Pfizer Inc., one of the world's largest biopharmaceutical companies, headquartered in Tadworth, Surrey. The UK team encompasses medical affairs, vaccine development, health economics, and evidence synthesis functions – all of which depend on systematic literature reviews (SLRs) to generate the clinical evidence that underpins regulatory submissions, market access dossiers, and public health research. Given the volume and complexity of SLRs conducted across therapy areas, any meaningful reduction in the time burden of evidence synthesis has direct implications for both research quality and speed to insight.

Authors: Jack Said, Charlie Reynard, Eleanor Olivia Fowlds (Pfizer Ltd., Tadworth, UK); Shayan Ali (DistillerSR, Ottawa, Canada) Study registered in PROSPERO: CRD42025635806