EMBEDDING AI-ASSISTED REVIEW OF UNSTRUCTURED CLINICAL DATA INTO ONCOLOGY TRIAL OPERATIONS

Nancy Bouvier[1]; Leemor Yuravlivker[1]; Rohan Singh[1]; Bo Young Kim[1]; Brian Liebe[1]; Michael Buckley[1]; Renata Panchal[1]

[1]Memorial Sloan Kettering Cancer Center

BACKGROUND

- Performance status (PS) is a required data element in oncology clinical trials and is commonly used for eligibility, treatment assessment, and safety reporting. In practice, PS is often documented only in unstructured clinical notes, requiring data managers (DMs) to manually search and review clinical documentation before entering data into sponsor electronic data capture (EDC) systems.
- This process is time consuming, error prone, and contributes to operational burden. Although artificial intelligence (AI) has potential to assist with unstructured data, many solutions fail to achieve sustained use due to poor workflow fit, limited transparency, and concerns about automation replacing DM judgment. To ground this work in a real data management workflow, we used performance status as the initial test case.

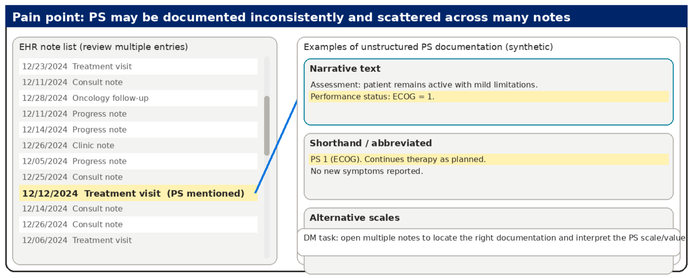

Figure 1. Pain point: PS is scattered across many notes and documented in inconsistent free text formats.

GOAL

The goal of this work was to develop and operationalize an AI-assisted feature that augments the review of unstructured clinical data within existing workflows, while preserving source transparency, regulatory accountability, and DM decision ownership.

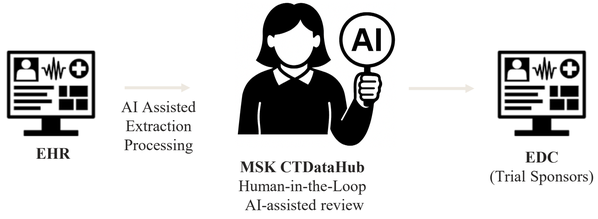

Figure 2. System context for AI-assisted review in MSK CTDataHub. Unstructured documentation from the EHR is processed to surface candidate values in CTDataHub; data managers review the supporting source context (human-in-the-loop) and retain decision ownership by confirming (or rejecting/editing) values before entry into sponsor EDC systems.

METHODS

- We implemented an AI-assisted review feature as a native capability within MSK's Clinical Trials Data Hub (MSK CTDataHub), a web-based application for DMs that surfaces ready-to-enter clinical trial data from source systems. PS data was selected as the initial use case; prior to this capability, MSK CTDataHub primarily supported structured data and retrieving PS required DMs to manually search unstructured notes in the EHR.
- The feature extends MSK CTDataHub by surfacing unstructured clinical information that requires review, specifically, AI-identified ECOG, KPS and Lansky PS scores documented within clinical notes. These values are presented to DMs for review and confirmation, rather than generated as free text output. It was designed using a human-centered, human-in-the-loop approach to support review and confirmation while preserving DM decision ownership. The AI-assisted review feature was deployed into production in October 2025 within a secure, on-premise environment.

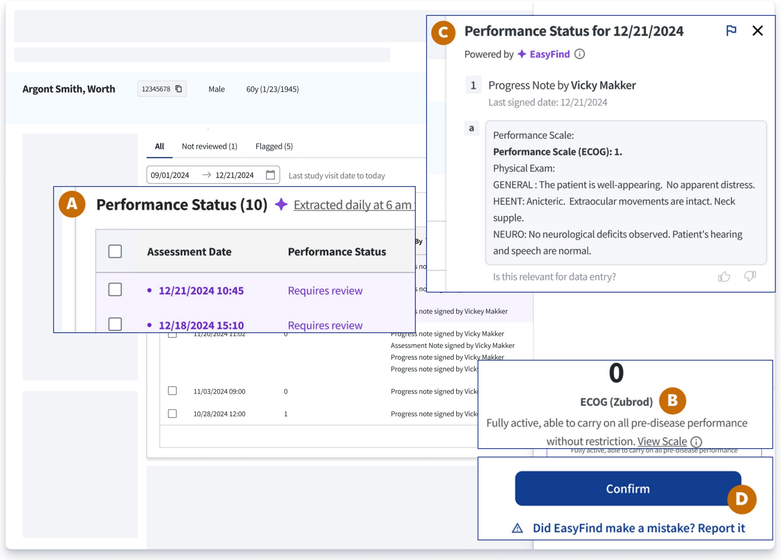

Figure 3. Confirmation workflow in MSK-CTDataHub (synthetic example; no PHI). (A) Unreviewed performance status (PS) candidates are surfaced in the Performance Status table. (B) The study required PS scale can be applied. (C) Selecting a row opens a side panel with extracted source note text for review. (D) The DM confirms the structured PS score (or reports an issue), preserving DM decision ownership and auditability.

RESULTS

- In the months following launch (October 2025 - April 2026), the feature was used across 23 oncology service groups by 225 DMs, with 21% adoption (47 active DM users) and 554 confirmed PS scores.
- Review without confirmation interactions were common during early adoption, which we hypothesize reflects DMs' exploratory use of the feature. During this initial evaluation period, no AI-identified performance status values presented for confirmation were corrected.

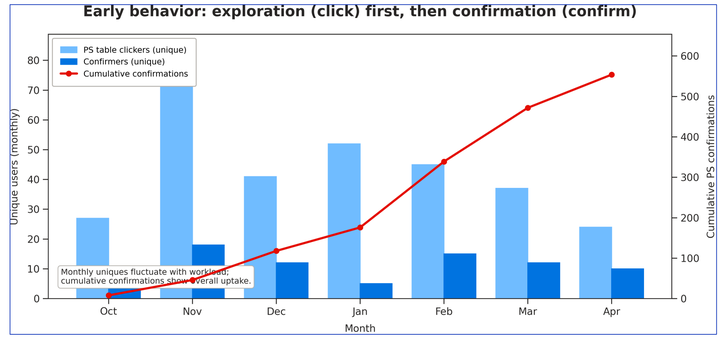

Figure 4. Heap-tracked interaction with PS review workflow in CTDataHub (Oct 2025–Apr 2026). Bars represent monthly unique users who clicked the PS table and monthly unique users who confirmed PS; the line represents cumulative PS confirmations over the same period.

23	225	47 (21%)	554	0
Service Groups	Data Managers	Active Data Managers	Confirmed PS	Corrections

CONCLUSION

- Our experience suggests that the AI-assisted approach is most effective when implemented as a review aid embedded within existing data management workflows rather than as an automated solution.
- Review without confirmation behavior highlighted the importance of change management, including clear workflow guidance and onboarding, to support user confidence while preserving data manager decision ownership.
- Future work will expand this feature to additional unstructured data elements and continue evaluation of user interaction to assess its impact on trial operations.