

Validation of an Agentic Large Language Model (LLM) System in the Extraction Stage of a Real-Time AI-assisted Living Systematic Literature Review (REAL-SLR): a Solution to Instant and Easy access to Clinical Trial Data (CTD)

Rozee Liu¹, Rhiannon Campden¹, Eddie Xiaole Liu², Triston Grayston³, Oscar Correa³, Anna Forsythe¹

¹Oncoscope-AI, Miami, FL, USA; ²Independent, Toronto, ON, Canada; ³Eviviz Inc., Vancouver, BC, Canada

CONCLUSIONS

- The agentic LLM system with a RAG architecture demonstrated high accuracy in extracting publications
- These findings suggest the system can enable real-time clinical data generation, supporting faster evidence development for HEOR decision-making and potentially improving patient access

BACKGROUND

- Health economics and outcomes research (HEOR) professionals face challenges staying current with clinical trial data
- De novo systematic literature reviews (SLRs) are often time-intensive with tight deadlines and large corpus of publication hits from searches.
- Data extraction of various types are required for SLRs.

OBJECTIVES

- To evaluate the feasibility of using an agentic large language model (LLM) for clinical trial data, assessing extraction accuracy and potential time savings with a Real-Time AI-assisted Living Systematic Literature Review (REAL-SLR), following Cochrane's initiative of Adherence to the Responsible AI use in evidence SynthEsis (RAISE)¹

Figure 1. Distribution of accuracy by PI/COS and accuracy threshold



Table 1. Annotation variables and accuracies by PI/COS

Population		Outcomes - Efficacy	
Clinical stage	99.00%	OS measure	95.95%
Histology	96.00%	Landmark survival	95.80%
Sub-population	95.00%	Median survival	83.60%
Biomarker	85.00%	Observation time	Iterating
Risk	83.00%	Primary progression measure	Iterating
Treatment path	Iterating	Median primary progression	Iterating
		Landmark primary progression	Iterating
Intervention/Comparator		Other progression measures	Iterating
Interventions	Iterating	Median other progression	Iterating
Intervention category	Iterating	Landmark other progression	Iterating
		Response measured	Iterating
Study design		Response data	Iterating
Randomization	100.00%	Outcome – QOL/Subgroup/Safety	
Phase	100.00%	QOL measure(s)	98.80%
Name	95.18%	Safety data	94.00%
Registration	100.00%	QOL data	90.36%
Follow-up	94.44%	Subgroups names	Iterating
N	90.36%	Subgroup data	Iterating
Analysis type	Iterating		

Abbreviations: N, number; OS, overall survival; QOL, quality of life

Table 2. Overall survival (OS) data extraction formats

Annotation variable	Annotation format			
	RCT	RCT, Cohorts	Non-RCT	Non-RCT, Cohorts
OS measure	Only these options are allowed. Only abstracts with actual data reported are annotated with "OS" below in this column. Those that only mentioned "OS" as an endpoint (or other forms of mentioning) without data reported should be "NR"			
OS median	"XX months vs. XX months; HR 0.XX; CI 0.XX-0.XX; p=0.XXX"	"Cohort 1: XX months vs. XX months; HR 0.XX; CI 0.XX-0.XX; p=0.XXX; Cohort 2: XX months vs. XX months; HR 0.XX; CI 0.XX-0.XX; p=0.XXX"	"XX months"	"XX months; XX months"
OS landmark	"XX% vs. XX%; p=0.XXX"	"Cohort 1: XX % vs. XX%; p=0.XXX; Cohort 2: XX % vs. XX%; p=0.XXX"	"XX%"	"XX months; XX months"

Abbreviations: CI, confidence interval; HR, hazard ratio; p, p-value; RCT, randomized controlled trial

RESULTS

- Our agentic LLM system generated annotations for 32 extraction variables for 3,098 (1,200 PC, 1,898 BC) publications
- Fourteen out of 32 variables achieved above 90% accuracy, 10/14 of which were above 95%. (Figure 1, Table 1)
- In addition to accuracy, our agentic system was able to produce annotations exactly as instructed
- As an example, overall survival (OS) extraction includes 3 variables: OS measure, median, landmark. Standardized annotation formats for these variables have been specified (Table 2)
- Only those matching the specified formats were considered a correct annotation. The accuracy of these 3 variables were 93.62%, 83.6%, and 90.9%

REFERENCES

1. Thomas J, Hair K, Noel-Storr, A. et al. Responsible use of AI in evidence SynthEsis (RAISE): recommendations for practice (version 3; updated 13 March 2026). In: Open Science Framework <https://osf.io/7j>. Washington DC: Center for Open Science. DOI: 10.17765/OSF:ID/PWAUD (accessed April 24 2026 via <https://osf.io/cqa8z>)

