Accelerating evidence synthesis in observational studies: An NLP-based solution for "living" systematic literature review

Jingcheng Du[1]; Frank J. Manion[1]; Dong Wang[2]; Long He[1]; Jingqi Wang[1]; Xiaoyan Wang[1]; Yeran Li[2]; David Eckels[2]; Peter C. Fiduccia[2]; Nicole Cossrow[2]; Lixia Yao[2]

[1]Melax Tech, Houston, TX, USA; [2]Merck & Co., Inc., Rahway, NJ, USA

Systematic literature reviews (SLR)

Systematic literature reviews (SLR), widely recognized as a robust method to identify and summarize evidence from published sources, are a critical tool in health economics and outcomes research (HEOR).

Benefits of SLRs

- Research knowledge base construction
  - Build a knowledge base to assist research gap analysis and direct research to support hypothesis
- Evidence synthesis
  - Use of statistical methods (ie, meta-analysis) to summarize the results of these studies
- Regulatory submission
  - FDA guidance requires surveys of the literature as sources of safety, efficacy, and "known uses" information

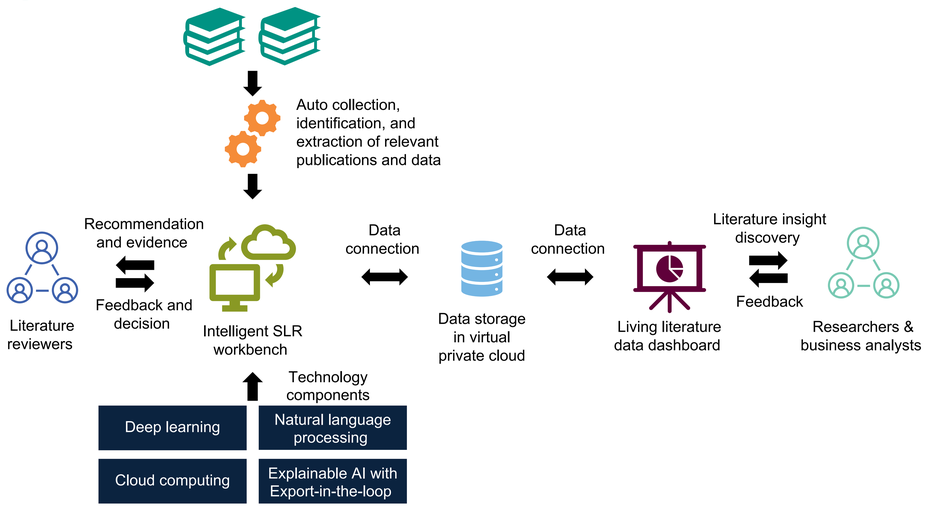
Challenges of SLRs

- Labor-intensive
  - Numerous relevant studies need manual screening and data extraction
- Error-prone
  - "To err is human". Tedious and repetitive work leads to unavoidable error
- Lack of traceability
  - Literature data are aggregated in Excel format and not easily linked to original evidence
- Not "living"
  - Challenging to keep up-to-date on the unprecedented growth of biomedical literature

Artificial intelligence-assisted literature analytics (AILA)

Artificial intelligence-assisted literature analytics (AILA) is a user-centered NLP-based software solution for SLR. AILA covers all major SLR steps and employs NLP algorithms to support the reduction of major time-consuming SLR steps, including abstract screening and data element extraction from full-text articles.

Figure 1.

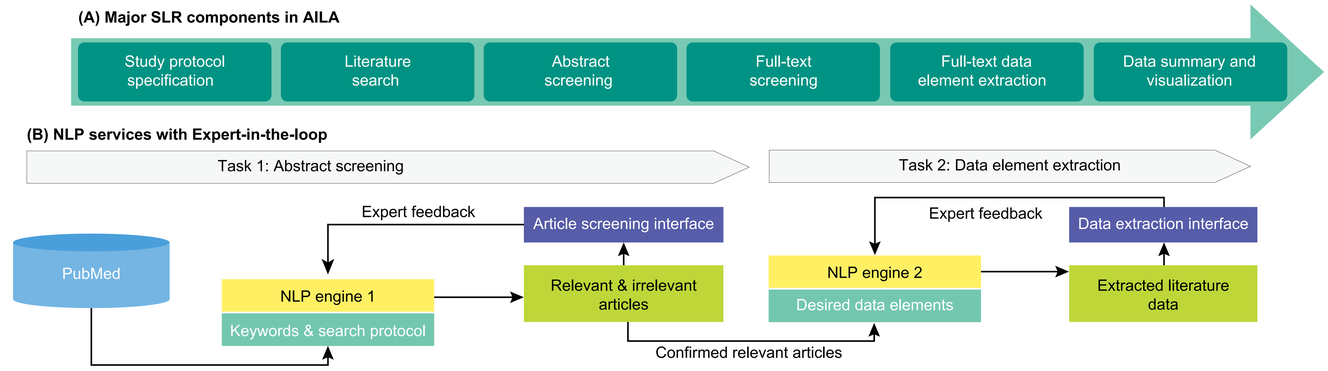


AILA consists of two major components that target two types of users within observational studies in an SLR life cycle: **(1)** An **intelligent SLR workbench** for literature reviewers who conduct routine literature reviews; **(2)** A **living literature data** dashboard for researchers and analysts who focus on analyzing SLR data and keep up-to-date on new evidence.

Natural language processing (NLP) in SLR

NLP refers to AI technologies that can extract structured information from textual data. AILA embeds NLP technologies for both abstract screening and data element extraction in SLR. For abstract screening, The system predicts articles' relevance status based on their title, abstract, and other metadata and highlights supporting information (eg, salient words that are impactful to inclusion and exclusion). For data element extraction, the system parses full-length articles and extracts data elements from both texts and tables. Human experts can further review and modify the predictions

Figure 2.



NLP validation

Table 1. Evaluation metrics for abstract screening

Task	Algorithm	F1 score	Precision	Recall	Accuracy
HPV prevalence (n=1,697)	XGBoost	0.808‡	0.769	0.851‡	0.888‡
	Support vector machine	0.727	0.781	0.681	0.859
	Logistics regression	0.684	0.897	0.553	0.859
	Random forest	0.523	0.944‡	0.362	0.818
Pneumococcal economic burden (n=421)	XGBoost	0.750‡	0.857‡	0.667‡	0.907‡
	Support vector machine	0.533	0.667	0.444	0.667
	Logistics regression	0.333	0.667	0.222	0.831
	Random forest	0.429	0.600	0.333	0.814
Pneumococcal epidemiology (n=211)	XGBoost	0.667‡	0.533	0.889‡	0.619
	Support vector machine	0.667‡	0.667‡	0.667	0.861‡
	Logistics regression	0.429	0.600	0.333	0.619
	Random forest	0.615	1.000	0.444	0.762

Table 2. Evaluation metrics for data element extraction

	HPV prevalence			Pneumococcal epidemiology			Pneumococcal economic burden		
	CRF	LSTM	Clinical BERT	CRF	LSTM	Clinical BERT	CRF	LSTM	Clinical BERT
Micro-average F-1 score	0.730	0.883‡	0.873	0.302	0.574	0.583‡	0.330	0.656‡	0.589
Macro-average F-1 score	0.6	0.790‡	0.739	0.358	0.578‡	0.519	0.311	0.541	0.566‡

Three previously completed SLRs were used to guide and validate NLP development, including: **(1)** the prevalence of human papillomavirus (HPV) detected in head and neck squamous cell carcinomas (referred to as HPV Prevalence); **(2)** the epidemiology of the pneumococcal disease (referred to as Pneumococcal Epidemiology); and **(3)** the economic burden of pneumococcal disease (referred to as Pneumococcal Economic Burden).

For abstract screening, the NLP model classifies relevance status for each article based on its title, abstract and other citation meta data such as journal, keywords, MeSH terms, article types, authors, etc. To build the abstract screening module, we evaluated several machine learning-based document classification algorithms.

For data element extraction, we treated the problem of data element recognition and extraction as a Named Entity Recognition (NER) problem, which aims to recognize the mentions of entities from the text. We evaluated a series of NLP algorithms consisting of ML and DL algorithms to recognize and extract data elements from full-text, including **1)** Conditional Random Fields (CRF), a classic statistical sequence modeling algorithm; **2)** Long Short-term Memory (LSTM), a variation of Recurrent Neural Networks (RNNs) and **3)** Clinical BERT, a novel Transformer-based deep learning model.

‡Highest performance.

Summary and Future Work

Incorporating state-of-the-art NLP algorithms, AILA aims to automate and expedite SLR process and facilitate "living" SLRs, thus allowing the system to proactively and continuously update relevant literature in a timely manner. This empowers scientists to dedicate more time to ensuring data quality and synthesizing evidence, while staying current with literature related to observational studies.

The system's generalizability to diverse SLR tasks across multiple therapeutic areas will be consistently improved and evaluated. Additionally, recent advancements in Large Language Models (LLMs), such as ChatGPT/GPT-4, may be integrated into the system to further enhance NLP accuracy and cater to the dynamic and varied needs of SLR tasks within the pharmaceutical industry.

Presented at ISPOR 2023; Boston, MA, USA; May 7-10, 2023.