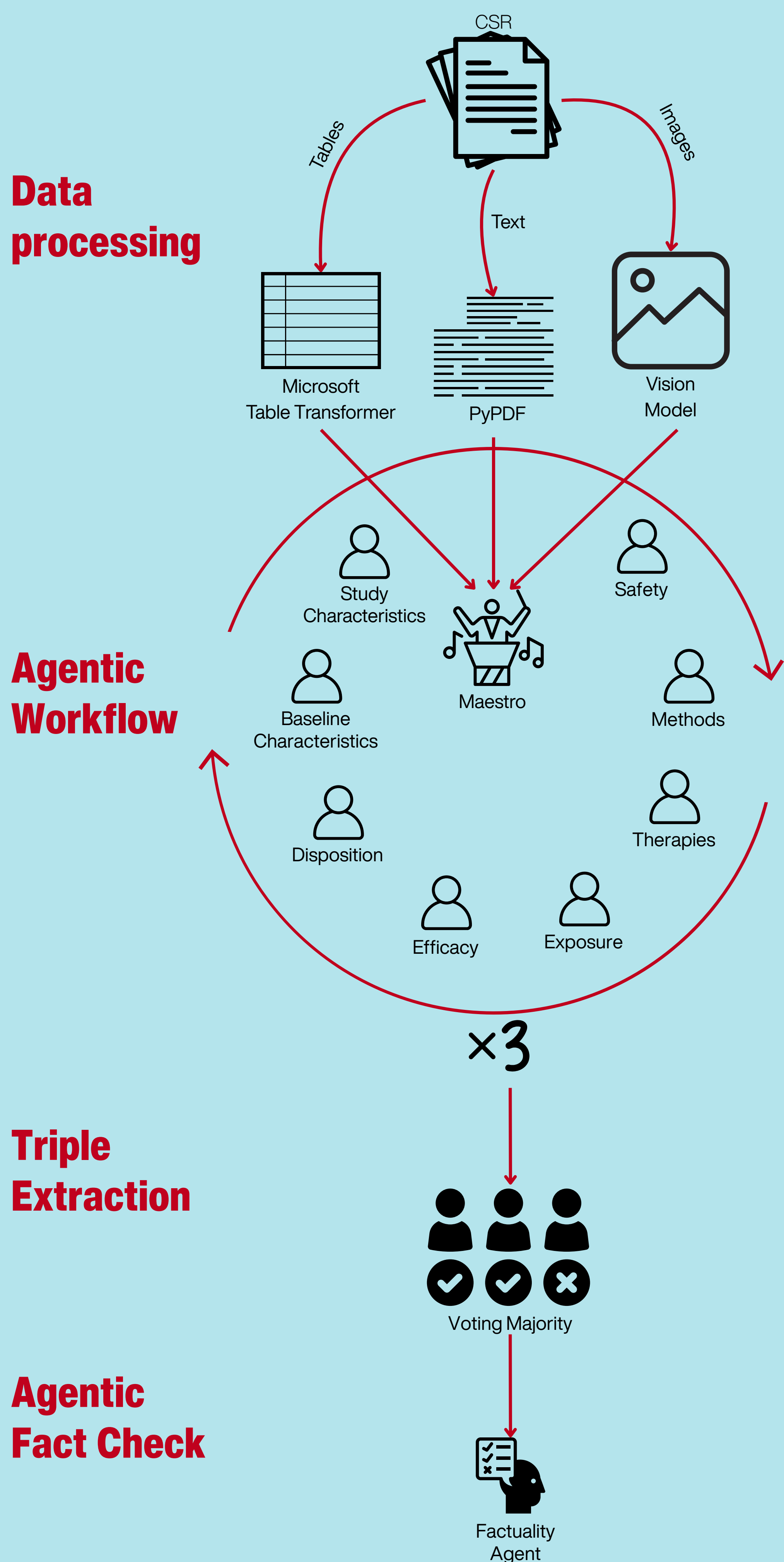


# Using Autonomous Generative AI Agents for Data Extraction from Clinical Study Reports: Accuracy Assessment Against Canada's Drug Agency Reports

#MSR9

Ghayath Janoudi, MBBS, MSc, PhD<sup>1</sup>, Mara Rada (Uzun), BA, MA<sup>1</sup>, Susan Mirabi, BSc, MSc<sup>2</sup>, Hussein El-Khechen, BSc, MSc<sup>2</sup>, Nicole Morse, BSc, MSc<sup>2</sup>, Andrea Lau, BSc, MSc<sup>2</sup>  
<sup>1</sup>Loon, Ottawa, ON, Canada; <sup>2</sup>PDCI, Ottawa, ON, Canada

## Agentic AI System Architecture



## Objectives

The manual extraction of data from Clinical Study Reports (CSRs) for evidence synthesis and health technology assessment is time-intensive and prone to error. This study evaluates the accuracy of data extraction performed by a system of multiple generative AI agents, compared with human-extracted data from a dupilumab reimbursement report by Canada's Drug Agency (CDA-AMC).

## Methods

Nine specialized AI agents were deployed on a proprietary multi-agent platform to extract study characteristics, baseline demographics, efficacy outcomes, and safety data from the publicly available sections of the Liberty AD PRESCHOOL CSR for dupilumab in pediatric patients with moderate to severe atopic dermatitis. A tenth AI agent validated the outputs from each extractor. These specialized agents were powered by various foundational large language models (e.g., Sonnet 3.5, Gemini 1.5, GPT-4). The CDA-AMC report for dupilumab served as the ground truth for assessing accuracy.

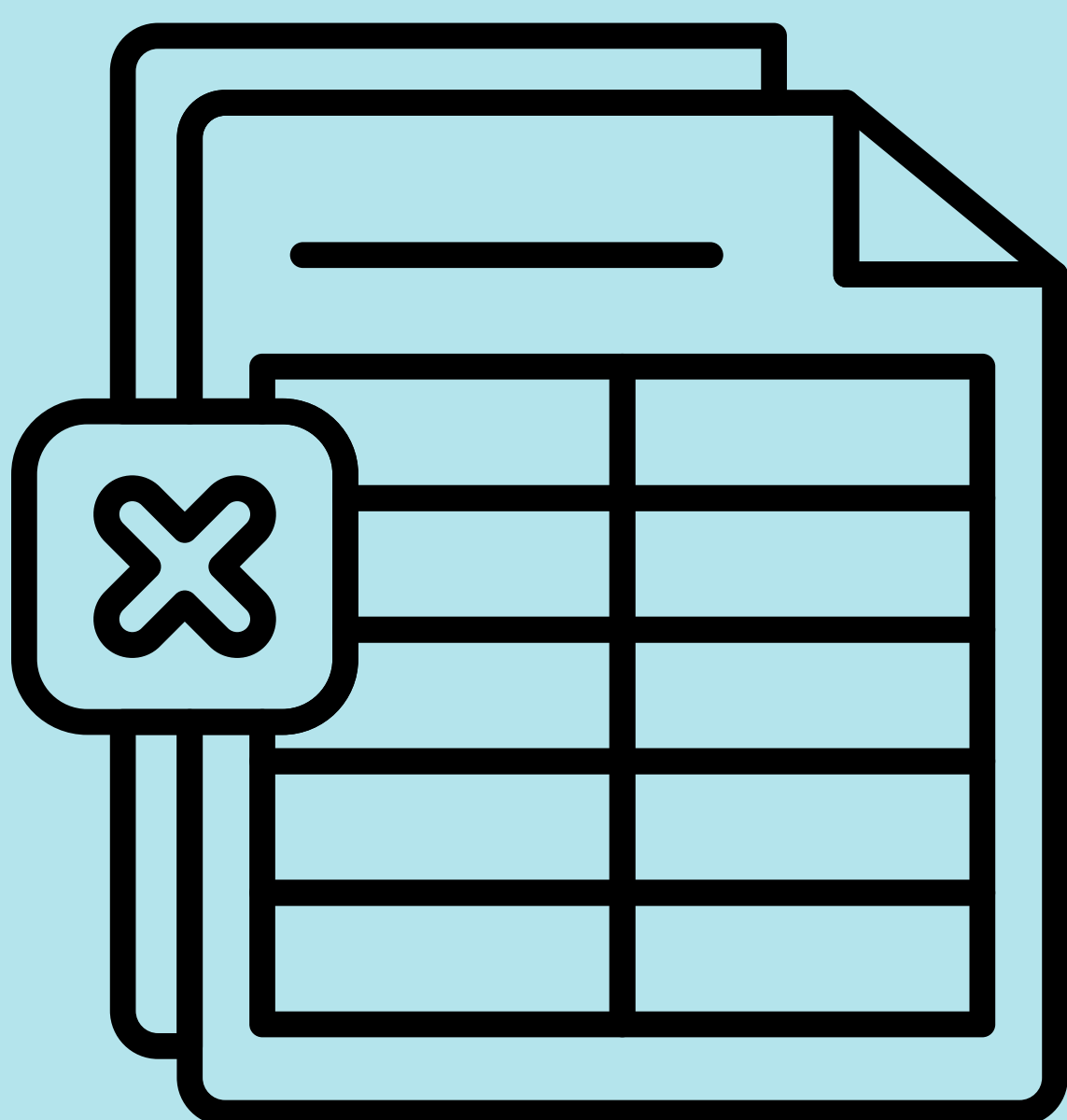
## Results

Out of 868 extracted data points, 793 were also reported in the CDA-AMC report. Of these 793 points, 777 (98.0%) matched the data in the CDA-AMC report, and 16 (2.0%) did not match. The remaining 75 data points were not available in the CDA-AMC report for direct comparison. Extracted data included point estimates, variability measures, and textual information regarding study design, statistical methods, patient disposition, treatment exposure, and efficacy and safety endpoints.

## Conclusions

A multi-agent generative AI approach for data extraction from a CSR demonstrated a 98.0% match with data in the CDA-AMC report. While these findings show promise for automating data extraction in health technology assessments, further research is needed to confirm reproducibility and assess performance across different study designs and therapeutic areas.

868 data points



Extracted Data

98.0 %  
Accuracy

95% CI:  
96.97% to 98.86%

793 data points



CDA-AMC Report